JOURNAL *of* EDUCATIONAL RESEARCH

Volume XXXIII MAY, 1940 Number 9

EVALUATING EVALUATION

PEDRO T. ORATA
Washington, D. C.

Editor's note: Much is being written these days about newer methods of measurement. The author in a stimulating paper calls attention to the shift in emphasis from measurement to evaluation and the assumptions underlying the new movement with critical comment. He concludes that evaluation, as now practiced, has a long way to go before it can justify association with "new school practices."

I. TOWARD *a New Philosophy of Measurement.* The change in terminology from "testing" and "measurement" to "appraisal" and "evaluation" is a significant development in education. The earlier view assumed that educational processes could be standardized and educational products "measured" by purely objective techniques. It was based on the mechanistic conception of the learning process, of which Thorndike is one of the strongest sponsors. Once standardization was accomplished, "guess work would be abolished, teaching would become certain and sure, and teachers 'scientific.' " (45)

In recent years, the trend has been to regard education more as a process of growth and development than as one of mere acquiring of skill and information and to emphasize the importance of attitudes and appreciation, interests, aims and purposes, emotional stability and control, personal and social adjustment, functional information, application of principles, interpretation of data, social sensitivity, study skills and work habits, creativeness, and like processes. (65, 82, 111, 114, 115, 116, 127)

This shift in emphasis is in part due to the change in the conception of learning from mechanism to organicism. On the technical side of measurement there has been a corresponding change toward the development of a broad philosophy of evaluation with emphasis not on the relative merits of the various forms of tests, but on the "kinds of evidence which indicate the attainment of various important outcomes of teaching." (99)

This change in point of view is expressed forcefully by McCall, who is undoubtedly one of the pioneers and popularizers of the Thorndike conception of measurement, in the 1939 revision of his well-known *How To*

Measure in Education, which contains a re-statement and elaboration of Thorndike's theses. McCall says: (51: vii)

Twenty years ago, I found measurement in danger of becoming, like medicine, the property of professional elite. * * * More and more in late years I have felt that measurement is in danger of becoming an ally of the *status quo.* It has tended to spin endlessly about itself a protective web of statistical intricacy, untroubled by any philosophical spark, and undisturbed by the world's travail. This book has been written to yank measurement out of its statistical complacency, infuse it a new spirit, sensitize it to the life that is outside as well as inside the textbooks, place it, as in former years, in the van of education.

Infuse measurement a "new spirit," "sensitize it to the life outside . . . the textbooks,"—that idea is a phenomenon and not just a change in point of view, as will presently be shown.

II. *The Functions of Evaluation.* The new philosophy of measurement is based upon an educational theory that has evolved from a number of hypotheses. "Among these are the belief that the classroom is a form of democratic social life by means of which children reconstruct their experiences, that these experiences grow from children's social activities which may be integrated around central problems suggested by the social activities, that pupils' interests and powers are developed by activities rather than by passive assimilation of knowledge, and that education is the foundation of social progress." (116: 212)

According to this point of view the functions of evaluation are no different from those of the school as a whole, namely: (1) to help provide more intelligent guidance of teaching and learning; (2) to develop more effective curricula and educative experience; (3) to secure more intelligent and effective coöperation with parents and community; and (4) to provide an adequate and objective basis for measuring, recording, and reporting progress (82) that facilitates rather than hindering the desired learning. (86)

III. *The Hypotheses of Evaluation.* Some years ago Thorndike announced a number of theses or hypotheses of measurement headed by the famous statement, "Everything that exists exists in some amount and can be measured." (51) The influence of this statement upon workers in the field of educational tests and measurements has been profound. The new point of view in evaluation has back of it a similar series of theses or hypotheses, which have been summarized by Wrightstone as follows: (116: 152)

1. The first hypothesis is that curricular change and evaluation are coördinate aspects of true experimentation. The end result of both is to

provide more insight and richness for the reconstruction of pupil personalities. The evaluation must, therefore, be compatible with the purposes toward which the curricular change is to contribute.

2. A second hypothesis is that a program of evaluation should be comprehensive. It should not be limited to a few isolated goals, or objectives, but should include all the major objectives of instruction. Evidence is available from a variety of sources to indicate that an index from one or two tests does not necessarily provide a measure of achievement for other major objectives of instruction.

3. A third hypothesis is that for a variety of major objectives of instruction no adequate methods or instruments for collecting reliable evidence are available. Until valid and reliable techniques are evolved, such objectives must be appraised by as careful subjective means as possible.

4. A fourth hypothesis is that a variety of means and techniques must be used for gathering evidence. These will range from pencil-and-paper tests, ratings, anecdotal records, and controlled observation techniques to physiological devices such as the psychogalvanometer. New techniques must be developed, and old techniques must be revised and modified to meet new needs.

5. A fifth hypothesis is that the measures will correspond to functional units of pupil behavior in actual curricular situations rather than to isolated units of behavior, often derived and defined from isolated clinical investigations. Too often the results of special clinical investigations have been transferred to the classroom and accepted as valid guides for influencing behavior.

6. A sixth hypothesis is that reliable and valid objective instruments of measurement are by their very nature restricted to an appraisal of limited aspects of pupil behavior. It is impossible to measure the whole result of an educative experience by any one test or battery of tests. The hope remains, however, that by measuring many important and vital aspects of educative experiences some valid appraisals may be obtained of the relative merits of diverse educational practises.

7. A seventh hypothesis, accepted by many educators, is that measures of functional behavior can best be developed by teachers working in coöperation with test technicians. In this conception the teacher and test technician will be partners in a coöperative enterprise.

IV. *Kinds of Changes in Pupils To Be Evaluated.* A comprehensive program of evaluation should give evidence not merely of the amount of information acquired by the pupils, but of the degree to which all of the objectives of the school program are being realized. What kinds of changes do we expect in boys and girls as a result of the educational opportunities that have been provided? Following are a few of the kinds for which evaluation has been attempted with some degree of success: (1) acquisition of functional information, (2) familiarity with technical

vocabulary, (3) familiarity with and facility in the use of important and dependable sources of information, (4) facility in gathering, organizing and interpreting data, (5) facility in formulating and applying significant principles to new problems or situations, (6) facility in planning and applying specific methods of investigation, (7) facility in the use of efficient and economical work skills and study habits, (8) interest in and proper attitude toward things, ideas, and people, (9) habit of working coöperatively with others, (10) habit of collecting and considering appropriate evidence and information before making important decisions, (11) leadership and organizing ability, and (12) social sensitivity. (97, 84, 85, 89, 86, 65, 66, 110, 127)

In view of these newer objectives of education the school subjects are expected to show correspondingly new kinds of results. *Art* should help develop initiative, interest, judgment, and coöperation; (26, 37); the *physical sciences*, the ability to use experimental methods in gathering, organizing, and interpreting scientific data and in applying scientific facts and principles (5, 10, 29, 31, 33, 34, 38, 54, 88, 89, 90, 129); *English literature*, the reading of literature understandingly, a broader understanding of life, greater sensitivity to social problems, and increasing intelligence with regard to human motives and purposes (24, 47, 77, 123); *French* and *Latin*, sufficient command of French and Latin vocabularies for simple reading and speaking (121, 122, 124); *home economics*, better health of and happier home life for all members of the family (12, 13, 14, 15, 46, 58, 59, 78); *mathematics*, thoroughness and precision in thought and action, disposition to question the validity of assumptions, expressed or implied, and sensitivity to the logic of arguments (27, 40, 41); *nursing*, proper attitudes toward patients, other nurses, and physicians and a wide range of interest, not only in nursing, but also in other and related fields, as well as the proper habits and skills in performance of nursing activities (19, 87); *social studies*, sensitivity to and disposition and ability to deal with social problems in an intelligent manner, interest in human welfare, and attitudes favorable to social improvement (4, 36, 43, 56, 68, 86, 105, 106, 112, 113, 117, 118); *health* and *physical education*, physical fitness (125).

Because of the changes in emphasis in the objectives of the school subjects, evaluation takes on a new meaning in supervision (100) and in the field of teaching methods, such, for instance, as in determining the effectiveness of motion pictures (32) and radio (17, 49, 55, 73, 74, 83, 104, 119, 120) as educational devices.

V. *Techniques of Constructing Achievement Examinations.* In the new conception of evaluation the following are basic: (1) the objectives of the total educational program, (2) the types of situations that normally evoke the kinds of experiences that the objectives call for, (3) the kinds of evidence in pupil behavior in these situations that are valid and reliable indications of change or progress toward the objectives, and (4) the instruments of evaluation with which to measure these changes. From these basic considerations follow the steps of constructing achievement examinations, to wit (95, 96, 98, 67):

1. Specifying the objectives of the school program as a whole in concrete and understandable terms so that the members of the teaching staff, the pupils, and their parents have a common meaning and interpretation of each item in the list.
2. Re-stating, if necessary, each of these objectives in the light of the nature, characteristics, and requirements of the course, field, unit, or area in the school program that is to be evaluated.
3. Defining the types of behavior that normally show whether or not and to what extent the objectives are being realized.
4. Selecting test situations that will evoke the types of student behavior patterns in such a way that they can be measured, recorded, and reported without hindering the desired learnings.
5. Trying out these test situations with a view to improving their validity and reliability and, at the same time, working toward making their use and administration more practicable.

VI. *Evaluation Instruments.* It goes without saying that a variety of new instruments is needed to measure what are generally regarded as intangibles in education. This is not to say that the objective or even the essay type of test has no place in the newer conception of evaluation. As a matter of fact, as will be shown presently, every effort is made to make use of objective examinations and other instruments in order to produce results that can be expressed in quantitative and precise form; that is to say, to give tangible evidence for the "intangible" objectives (128). In addition to these instruments, there are being developed new techniques of appraisal such as so-called "scales of beliefs," "interest questionnaires," "nature of proof tests," "application of principles tests," and the like, which differ, not in form but in content, from the well established measuring instruments. Most of the new instruments, however, are flexible as to time allowed for them, and some of them do not have any time limit at all.

Aside from tests and scales, there are other instruments of evaluation that are not ordinarily regarded as such, as, for example, anecdotal rec-

ords, data from questionnaires, records of attendance, records of physical examinations, diaries kept by teachers and pupils and parents, case study records, and other kinds of records that show or give evidence of significant changes in the behavior and development of the boys and girls in school. They may be formal or informal records, complete or incomplete records; just as long as they give bits of significant and valid evidence of changes in the pupils they are properly regarded as evaluation instruments. What seems important is not that each instrument is complete in and of itself, but that it contributes, along with other instruments, some evidence, slight though it may be, toward a comprehensive and dependable evaluation of the total school program.

Following are a few of the evaluation instruments developed by the staff of the Evaluation in the Eight Year Study under the direction of R. W. Tyler.[1]

Test 4.2. A Scale of Beliefs (P. E. A. 2519, 1937). This test is intended to measure or give evidence concerning (1) the points of view or attitudes held by students within the areas of militarism, nationalism, racialism, democracy, economic individualism, and labor and employment; (2) the strength or intensity of these attitudes; and (3) the conflicting points of view or inconsistencies present. It consists of a series of statements representing a variety of beliefs and attitudes. (25)

Test 2.5. Interpretation of Data (P. E. A. 2761, 1938). Time allowed—45 to 60 minutes. In this test the student is given the opportunity to judge certain interpretations of meanings that have been derived from charts, tables, and prose data relating to various issues in social studies and science. Ten problems are given, under each of which are a number of statements that might be called "interpretations" or "inferences" drawn from the facts. The student is to judge these statements according to the following standards: (1) fully supported by the facts, (2) probably true, but not certain, (3) facts insufficient to warrant a positive or negative statement, (4) the facts suggest that the statement is probably false, and (5) the facts contradict so that the statement is definitely false.

Test 2.51. (1939). Time allowed—about 90 minutes. Constitutes the second form of the original test prepared in 1938.

Test 7.1. Familiarity with Sources of Information (1938). Time allowed — 45 to 60 minutes. This is a test to see how familiar the student is with the more general sources of information which are frequently used by high school students. There are 40 problems, and 5 different possible sources of information listed under each problem. The student is to judge each one of the sources according to three standards as follows: (1) good place to go for information on the question, (2) it might have some helpful information, but not a very good place to go, and (3) not worth using in studying the question.

[1] Additional instruments are listed in the bibliography.

Test 1.3. Application of Principles — Aspects of Thinking (P. E. A. 2724, 1937). Time allowed — 45 to 60 minutes. This test has been devised to obtain evidence relating to the ability of students to apply certain principles commonly developed in science classes. The student is confronted with 10 problems, which supposedly have not been previously discussed in the regular classroom. He is to indicate which of the indicated possible conclusions and reasons he would be willing to accept.

Test 1.31. Application of Principles in Chemistry (P. E. A. 2679, 1938). Time allowed — 45 to 60 minutes. Twelve problems are given, and the student is given instructions similar to those for Test 1.3.

Test 1.32. Application of Principles in Physics (P. E. A. 2692, 1938). Time allowed — 75 to 90 minutes. Eighteen problems are given, and the student is given instructions similar to those for Tests 1.3 and 1.31.

Interest Index (8.2a — 1939). A list of 200 things to do, some of which appeal to a great many students and some of which appeal to only a few, such as, "to write stories," "to play basketball," "to learn to go about getting a job," etc. The student is to indicate his reaction to each item whether he likes or wants to do what the item says (L), neither likes nor dislikes it (I), or dislikes it or would not enjoy doing it (D).

Test 5.1. Problems Relating to Proof in Mathematics (P. E. A. 2763, 1937). Time allowed — 45 to 60 minutes. This test is composed of 13 reasoning problems based upon questions relating to public utilities, unemployment, democratic government, and similar topics. No special knowledge of these topics is required; all of the essential data are given in the situation. The student is to indicate which of the reasons and conclusions given he is willing to accept as his own. This test is similar in form to Test 1.3, *Application of Principles.*

Test 3.4. Literary Information Test — English Literature (1939). Time allowed is not indicated. This test consists of 14 exercises, each of which requires the student to identify certain literary items, such as titles by the name of the author, quotations by the name of the author, authors by types of writing, biographical information, literary contribution by author, authors by country in which they were born, characteristics of literary work by authors, characters by title of work, titles by literary period, titles by literary type, literary content by title of work, titles by type of subject matter, titles by setting of narrative, and titles by literary form.

Test 3.5. Literary Information Test — American Literature (P. E. A. 1291, 1937). Time allowed is not indicated. Similar in form to Test 3.4, but contains 16 instead of 14 exercises.

Test 1.4. Social Problems (P. E. A. 2719, 1938). This test deals with the disposition of students to apply various types of social principles and values to contemporary social situations. Some of these principles express concern for human and democratic values, and others defend the interests of privileged groups. The purpose of this test is to get evidence on various aspects of thinking in situations in which values and beliefs have a strong,

if not a determining, effect on thinking. There are six exercises in the test, dealing with the following problems: the effects of the use of labor-saving machines, housing, taxation, racial discrimination, freedom of speech, and health security of workers in factories. In each exercise the students are asked to choose from the listed courses of action those which, to them, seem most adequate and then support their choices by checking certain reasons from the list given in the test.

Test 3.6. A Questionnaire on Reading Interests and Reading Outcomes (P. E. A. 1935, 1937). The purpose of this questionnaire is twofold: (1) to discover what interests the pupils have developed as a result of their experiences in reading books, and (2) to discover what influences books have had on the pupils' insight into certain life problems. The pupil is to react to 35 questions by indicating the answer on the dotted line or checking one of the following: Yes, No, Uncertain, as the case may be. Part I aims to find out the interest developed in reading books, and Part II with the influence the books might have had upon the pupil. No time limit is indicated.

Test 3.7. Critical Mindedness in the Reading of Fiction. (P. E. A. 2615-a, 1938). This test is designed to ascertain whether the pupil is able to distinguish between fiction that is reasonably true to life and fiction that is not. Two short stories are reproduced for the student to read. Under each story are ten statements to which the pupil is to react by indicating whether he agrees with the statement (A), disagrees with it (D), or is unable to say (U). No time limit is indicated.

Test 3.8. Judging the Effectiveness of Written Composition (P. E. A. 2649, 1938). The following instructions to the pupil explain the nature and requirement of the test: "Read the following story, noting carefully the way in which the writer has developed the narrative. Then turn to the criticisms that follow it and indicate, in accordance with the directions, whether you agree or disagree with these criticisms." Twenty-eight statements follow the story, such as "The writer should not have included so many different episodes in one brief story," "*Prissey* is a good name for Esmond's wife," etc. No time limit is indicated.

Test 3.3. Questionnaire on Voluntary Reading (P. E. A. 2662, 1938). Time allowed — 40 minutes. In this questionnaire on voluntary reading, the student is asked to answer fifty-two questions with YES, NO, or NEITHER YES NOR NO. They are of many kinds and are classified into seven general types according as they indicate that the pupil (1) likes to read, (2) wants more reading, (3) is curious about his reading, (4) expresses appreciation in other media, (5) identifies himself with his reading, (6) relates his reading to his life, and (7) evaluates his reading.

Descriptive Trait Profile (P. E. A. 2655). No date or time limit is indicated. A scale of 14 traits, cheerfulness, contentment, poise, adaptability, self-confidence, friendliness, leadership, attitude toward the opposite sex and toward work and play, coöperativeness, discrimination, sensitivity to the rights of other people, and general maturity. The teacher is to

rate each student in terms of the following standards: *extreme, average, changeable.* In case of doubt the teacher checks the column under "Uncertain." Each trait is described briefly under each one of the three columns.

Evaluation of Reading (P. E. A. 90, Bulletin No. 4, January 1936). Contains two record forms, one for students, "Student's Monthly Reading Record," and another for teachers, "Form for Teacher's Monthly Record of Student Reading," and suggestions as to ways of securing, interpreting, and using the information obtained.

Checklist of Magazines. No date or number is given. Contains a list of 567 magazines. The student is to check those that were read during the three months preceding, how often and how much of each was read, and where obtained. The student is also asked to answer nine questions about his reading habits and attitudes.

VII. *Developing a Comprehensive Program of Evaluation.* One of the hypotheses of evaluation mentioned early in the report is that appraisal to be effective should be comprehensive in scope so as to include all the major objectives of the school program. It is on the basis of this hypothesis, primarily, that the evaluation program of the Thirty Schools in the Eight Year Study of Progressive Education Association is being developed. It is also on the basis of this assumption that Wrightstone's several appraisal studies (29) were formulated. While there are undoubtedly many difficulties that seem to limit what can be done, still ways and means of overcoming such difficulties have been suggested (86).

1. It is not necessary to set aside time for evaluation. Teaching and learning on one hand and evaluation on the other take place at the same time, and for this reason, "the record of the pupil's activities in learning, the products of his activities, the results of his own efforts, and experiences, are frequently records for appraisal."

2. The pupils, themselves, can aid materially in making appraisal of their own progress if they are made to realize the significance of evaluation for them and are encouraged to make self-appraisals, and if properly managed and motivated an evaluation program of the kind suggested can be organized without putting an undue burden upon the teachers or the pupils.

3. Economy of time and effort is possible by utilizing situations for making appraisals in more different directions than may be apparent on the surface. "For example, an inventory of the resources and needs of the local community by a seventh-grade pupil threw light upon his attitude toward the functions of social and economic processes, his ability to use certain types of social science information, his ability to understand and to use some technical terms in the social studies field, his facility in written expression, his ability to construct and to interpret certain kinds of maps and charts."

4. It is possible to develop indexes of pupil progress in those aspects

of his growth that are difficult to appraise directly. Temperature is measured indirectly, and so are many phases of physical and social phenomena.

VIII. *Critical Appraisal of the "New" Evaluation Movement.* The term "measurement" implies refined, exact, quantitative, and standardized assessments of *things* or *states of being* in an impersonal and objective manner. On the other hand, the terms "appraisal" and "evaluation" are usually associated with *values,* with *processes of growth and development* toward a goal that does not remain fixed. Measurement places emphasis on accuracy and reliability; evaluation, on validity and social utility. Measurement deals primarily with subject matter attainment and information; evaluation, with experience and motivation. Theoretically, evaluation has the upper hand, but in practise its friends and critics praise and condemn it on the same grounds, as the following comments tend to show:

Pro

C. M. Pruitt, commenting on Tyler's *Constructing Achievement Tests,* said: "The theses relating to ability to use the scientific method, measuring the ability to infer, what is statistical significance, permanence of learning, and techniques of evaluating behavior, are especially commendable." (11: 264.)

D. E. Scates commented on the work of Tyler and his staff in connection with the Evaluation in the Eight Year Study of Progressive Education Association as follows: "The work of the Evaluation Staff has been outstanding for the sincerity, the courage, and the resourcefulness with which intangible objectives have been defined and means of measuring them sought. The evaluation study may well prove to mark a turning point in the history of educational measurement." (75: 536.)

A. G. Schmidt's comment on Wrightstone's *Appraisal of Newer Practices in Selected Public Schools* is this: "The Study goes far toward satisfying a need which has been long felt for a scientific appraisal of the results of the new methods." (75: 272.)

Con

S. R. Powers says about the same book by Wrightstone: "The techniques which are described in some detail seem to be valid but not adequate." (75: 272.)

P. R. Pierce adds: "Failure to supply clearcut data regarding the practices evaluated seriously weakens the scientific validity of the study." (75: 272.)

W. J. Osburn corroborates thus: "The study is subject to the usual defects of lack of reliability, validity, and adequate sampling." (75: 272.)

Hilda Taba, commenting on Wrightstone's *Appraisal of Experimental High School Practices,* writes: "He (Wrightstone) has pointed out which differences are statistically significant, but has failed to inform the reader of the social significance of these same differences • • • Statistical sig-

nificance alone does not prove educational significance and since no norms are given the reader is at loss to determine the practical value of the results of this study." (75: 271.)

It would seem from the above comments by its friends and critics that the new movement in evaluation suffers from very much the same sort of shortcomings as does the type that it seeks to replace. Measurement has been criticized for its exclusive use of paper-pencil tests; the new tests of Tyler and Wrightstone and their staffs are mostly of the paper-pencil variety.[2] Measurement has often been condemned for the lack of validity of its evidence; evaluation does not seem to offer a much better and more valid measure of pupil progress,[3] and besides it suffers from lack of reliability quite as much as did the old type of essay examination.[4] Measurement has been criticized for its emphasis on standardized patterns or answers; evaluation has its own patterns and answers.[5] Measurement stresses quantitative assessment of the qualitative aspects of behavior; evaluation does nothing different. The fact that evaluation is often less exact and precise may only mean that it is more vague and confused than measurement. Vagueness and equivocation are no virtues, even in a confused world.

In final analysis, the present reviewer is of the opinion that many of the tests and evaluation instruments that are being used in progressive schools lack the qualities that would render them more appropriate for measuring educational progress than the ordinary objective tests—that, therefore, to say it bluntly, while teaching is apparently becoming more progressive, testing remains basically traditional. What has happened is that the old testing instruments have been renamed to fit the new educational program.

In fairness to all concerned, it should be granted that evaluation has made its most lasting and significant contribution, first, in its insisting upon evidence for what are generally regarded as intangibles in education; second, in its theory that testing is an integral part of the educative process; third, in its exploration of new sources of evidence; and, finally,

[2] With two exceptions the instruments described in Section VI are of the paper-pencil variety. (The instruments were furnished on request by Dr. Tyler's office.)

[3] With the two exceptions the evidence from said instruments does not go beyond words that indicate the students' knowledge about or of the situations provided, as in ordinary objective tests.

[4] The Descriptive Trait Profile, which is one of the exceptions, is general and impressionistic and, therefore, very highly subjective. The other exception, Evaluation of Reading, is equally unsatisfactory from the standpoint of validity of evidence.

[5] Twelve of the instruments provide or suggest the answers in the same manner as that in which ordinary objective tests do.

in its suggesting a plan for comprehensive evaluation. The fact that many of its tests are basically traditional does not lessen the value of its progressive outlook.

What seems needed now, theoretically speaking, is a functional combination of the outlook of evaluation and the method of scientific measurement conceived in a much broader sense than is apparent in ordinary educational tests and experimentation (21). Fortunately, as Hosic points out, "There is a fair prospect that the two (movements) will get together. No amount of science can ever enable us to dispense with human judgment. Measuring samples of behavior can furnish data, but not ready-made conclusions. Measurement is, therefore, only an accessory. The essential act is that of evaluation." (45: 448.)

IX. *Concluding Statement.* It seems evident that, contrary to what is generally believed, evaluation as actually applied by Tyler and Wrightstone has a long, very long, way to go before it can fully justify its being associated with "newer school practices." In theory, it must be admitted, there is agreement between Tyler's staff and the staffs of the thirty schools; in practise, however, there is considerable room for honest doubt. Tyler once said that the philosophers can say all they want to about what education is for, but after all, teachers teach for what is in the tests; and he concluded, "Show me who make the tests, and I'll show you who will remake the curriculum." One may well wonder what is happening in the Thirty Schools. Tyler and his staff make the tests in Chicago to be used or copied in Denver and New York. Teaching is done in the Bronx and testing on the Midway. A principal of one of the thirty schools, upon being requested for material on evaluation, answered thus: "We do the teaching. As for evaluating it, we leave it all to Dr. Tyler and his enterprising staff."

If evaluation is to be really progressive it should be, in fact, an integral part of the total educative process. A paper-pencil type of test made in advance by some-one on the outside for some-one else to use or to copy is inappropriate for the kind of program for which the newer schools stand. Evaluation, to be really an integral part of the educative process, should grow out of or emerge from that process. No-one, not even an expert, should evaluate the work of someone else, if the philosophy of progressive education is accepted as a point of orientation. In fact, the teacher should not evaluate the work of the pupils; they should do the evaluating themselves. If not, why emphasize the objective of self-direction, which obviously includes the process of evaluation of one's performance in the

light of the goal that had been self-determined? The evaluation function of the teacher is to provide situations whereby the pupils are led to evaluate their own procedures in the light of criteria that they, themselves, had formulated. In like manner, the function of an expert in evaluation, if there be one, is not to make tests for teachers to use or to copy, but to guide them in integrating their evaluation function with their teaching procedures.

When Raths made the trip to the Lakeshore School he used tests that were made in Columbus (66). The teachers looked on as the pupils took the tests, and when the results were tabulated it was found that they correlated with the teachers' judgment. There is nothing new in that. When Wrightstone compared the standard and activity schools, he found that the pupils in the activity schools did better in the objectives that were emphasized in those schools, and, unsurprisingly enough, the standard schools did better, generally speaking, along the lines that were stressed there (111, 112, 114, 115, 116). There is nothing startling about that, either. It only proves that education is possible. The situation reminds one of the generalization arrived at by two specialists who made a five-year experimental study on the subject of reading, namely: "Pupils will learn to read if they are taught to read." Well, why not?

The most immediate need is, therefore, first, for progressive education to look at its program of evaluation to determine its consistency with the objectives that are so insistently emphasized and, second, for the evaluation practises to be evaluated and subsequently modified so as to provide, in fact as well as in theory, "new tests for new needs." (110.)

Bibliography on the Newer Conceptions of Evaluation

1. Aikin, Wilford M. "The Purposes of the Eight Year Experimental Study," *Educational Record*, XVI (January 1935), pp. 107-120.
2. Alschuler, Rose H., and Hattwick, Laberta Weiss. "Measuring Social and Emotional Development." *Appraising the Elementary-School Program*. Sixteenth Yearbook (Washington, D. C.: Department of Elementary School Principals, National Education Association, 1937), pp. 502-507.
3. Arnold, D. L. "Parents and the Purposes of the School," *Educational Research Bulletin*, XVI (September 15, 1937), pp. 156-160.
4. Arnold, D. L. "Testing Ability To Use Data in the Fifth and Sixth Grades," *Educational Research Bulletin*, XVII (December 7, 1938), pp. 255-259.
5. Barrows, W. M., and others. "Coöperative Zoology Test." Provisional Form 1936. Coöperative Test Service.

6. Barthelmess, Harriet M. "The Measurement of Outlining Skills." *Reconstructing Education Thru Research.* Official Report of 1936 Meeting (Washington, D. C.: American Educational Research Association, a department of the National Education Association, 1936), pp. 260-263.
7. Beers, Fred S. "The Human Side of This Testing Business," *Educational Record,* XVII (October 1936), pp. 577-608.
8. Bergstresser, J. L. "Evaluation at the College Level in the Eight-Year Study of the Progressive Education Association," American Association of Collegiate Registrars Bulletin 12: (July 1937), pp. 235-242.
9. Brown, Marian, and Martin V. "Records as Tools in the Study of Adolescents," *Educational Research Bulletin,* XV (November 11, 1936), pp. 207-215.
10. Buckingham, Guy E., and Lee, R. E. "A Technique for Testing Unified Concepts in Science," *Journal of Educational Research,* XXX (September 1936), pp. 20-27.
11. Buros, Oscar K. *The Nineteen Thirty-Eight Mental Measurement Yearbook.* (New Brunswick, New Jersey: Rutgers University Press, 1938), 415 p.
12. Chadderdon, Hester. "Measurement in Home Economics Education," *Practical Home Economics,* XIII (October 1935), pp. 289; 308-309.
13. Chadderdon, Hester. "Planning a Measurement Program," *Practical Home Economics,* XIII (November 1935), pp. 322; 341-342.
14. Chadderdon, Hester. "Evaluation of Evidence in Measurement," *Practical Home Economics,* XIII, pp. 363; 374-375.
15. Chadderdon, Hester. *Problems Relating to Home and Family Life in Need of Research: An Examination of the Emphasis Given to Home and Family Life Under the Purnell Act.* Columbus, Ohio: The Ohio State University, (Ph.D. Thesis) 1938. 243 p. (Ms.)
16. Connor, William L. "Testing Techniques Adjusted to the Aims of Education for Character and Personality." *Education for Dynamic Citizenship.* Twenty-Fourth Annual Schoolmen's Week Proceedings (Philadelphia: School of Education, University of Pennsylvania, 1937), pp. 290-294.
17. Contant, F. R. "Determining the Appeal of Special Features of a Radio Program." *Journal of Applied Psychology,* XIII (February 1939), pp. 54-57.
18. Cowell, C. C. "Play Behavior and Personality Analysis," *Educational Research Bulletin,* XVI (October 20, 1937), pp. 182-186.
19. Crist, Alice L. "The Ohio Coöperative Testing Program in Nursing Education," *Educational Research Bulletin,* XVIII (February 1, 1939), pp. 31-39.
20. Culley, Benjamin H. "Measuring the Ability to Interpret Data." *California Journal of Secondary Education,* XIII (March 1938), pp. 152-154.

21. Dewey, John. *The Sources of a Science of Education.* (New York: Horace Liveright, 1929), 79 p.
22. Diederich, P. B. "Evaluation Records," *Educational Method,* XV (May 1936), pp. 432-440.
23. Diederich, P. B. "A Master List of Pupil Activities," *Educational Research Bulletin,* XV (September 16, 1936), pp. 166-169.
24. Eberhart, W. "The Teaching of Literature: An Approach to Evaluation," *Educational Research Bulletin,* XVII (January 19, 1938), pp. 1-6.
25. Eginton, Daniel P. "Tests and Measures of Satisfactory Growth," *Educational Method,* XV (December 1935), pp. 145-149.
26. Ellingson, Mark. "Evaluation of Art Accomplishment," *Role of Research in Educational Progress.* Official Report of 1937 Meeting (Washington, D. C.: American Educational Research Association, a department of the National Education Association, 1937), pp. 170-175.
27. Fawcett, H. P. *The Nature of Proof: A Description and Evaluation of Certain Procedures Used in a Senior High School To Develop an Understanding of the Nature of Proof.* Thirteenth Yearbook of the National Council of Teachers of Mathematics, New York: Teachers College, 1938, 146 p.
28. Fowler, Burton P. "Some Fundamentals in Evaluating Educational Programs," *Educational Record,* XVII (October 1936), pp. 23-34.
29. Frutchey, Fred P. "A Coöperative Program for Developing Tests of the Ability To Use Scientific Method in College Sciences," *Science Education,* XXII (February 1938), pp. 81-85.
30. Frutchey, F. P., and Hendricks, B. C. "Constructive and Validative Examinations," *Journal of Chemical Education,* XV (January 1938), pp. 40-43.
31. Frutchey, F. P. "Testing for Application of Scientific Method," *Educational Method* XV (May 1936), pp. 427-432.
32. Frutchey, F. P., and Dale, Edgar. "Evaluating Motion Pictures," *Educational Research Bulletin,* XV (September 16, 1936), pp. 163-165.
33. Frutchey, F. P. "Evaluating Chemistry Instruction," *Educational Research Bulletin,* XVI (January 13, 1937), pp. 1-6; 28.
34. Frutchey, F. P. "Illustrative Test Exercises in High School Chemistry," *Educational Research Bulletin,* XVI (May 19, 1937), pp. 122-126.
35. Ginsburg, Morris B. "Constructing and Using a Measure of Good Manners." *Appraising the Elementary-School Program.* Sixteenth Yearbook (Washington, D. C.: Department of Elementary School Principals, National Education Association, 1937), pp. 514-521.
36. Grim, P. R. "A Technique for the Measurement of Attitudes in the

Social Studies," *Educational Research Bulletin,* XV (April 15, 1936), pp. 95-104.

37. Grimes, J. W., and Bordin, E. "A Proposed Technique for Certain Evaluations in Art," *Educational Research Bulletin,* XVII (January 4, 1939), pp. 1-5.
38. Handorf, B. H. Coöperative Chemistry Test. Provisional Form, 1934-1937. Coöperative Test Service.
39. Harden, M. "Evaluating a School Program," *Teachers College Record,* XL (February 1939), pp. 391-406.
40. Hartung, M. L. "Some Problems in Evaluation." *Mathematics Teacher,* XXXI (April 1938), pp. 175-182.
41. Hartung, M. L. "Evaluating Appreciation of the Cultural Values of Mathematics," *School Science and Mathematics,* XXXVII (February 1937), pp. 168-181.
42. Hawkes, H. E., Lindquist, E. F., and Mann, C. R. (Editors). *The Construction and Use of Achievement Examinations.* (Boston: Houghton Mifflin Company, 1936), 497 p.
43. Herrick, J. H. "The Evaluation of Certain Aspects of Thinking in the Social Studies," *Educational Method,* XV, pp. 422-426.
44. Hopkins, Thomas. "The Problem of Measurement Involved in Integration." *Reconstructing Education Thru Research.* Official Report of 1936 Meeting (Washington, D. C.: American Educational Research Association, a department of the National Education Association, 1936), pp. 201-205.
45. Hosic, J. F. "The Problem of Evaluation." *Educational Method,* XV (May 1936), p. 448.
46. Huston, Hazel H. "Measuring Achievement in Home Economics," *Journal of Home Economics,* XIX (January 1937), pp. 19-22.
47. Jones, C. C. "Evaluating Achievement in Literature," *Educational Method,* XV (May 1936), pp. 416-421.
48. Lindquist, E. F. "Changing Values in Educational Measurement." *Educational Record,* XVII (October 1936), pp. 64-81.
49. Longstaff, H. P. "A Method for Determining the Entertainment Value of Radio Programs," *Journal of Applied Psychology,* XXIII (February 1939), pp. 46-54.
50. McCall, William A., Herring, John P., and Loftus, John J. "Measuring Achievement in Activity and Control Schools in New York City," *Teachers College Record,* XXXIX (February 1938), pp. 423-431.
51. McCall, W. A. *Measurement.* (New York, Macmillan, 1939), 535 p.
52. McDowell, John G., and Anderson, Howard R. "Testing the Ability of Pupils to Outline," *School Review,* XLVI (January 1938), pp. 48-56.
53. McGaughy, James R. *An Evaluation of the Elementary School.* (Indianapolis: Bobbs-Merrill Company, 1937), 421 p.
54. Noll, Victor H. "Measuring the Scientific Attitude," *Journal of*

Abnormal and Social Psychology, XXX (July 1935), pp. 145-154.

55. Ollry, Francis, and Smith, Elias. "An Index of 'Radio-Mindedness' and Some Applications," *Journal of Applied Psychology*, XXIII, pp. 8-18.
56. Orata, Pedro T. "Evaluation in the Field of Social Science," *Educational Method*, XVI (December 1936), pp. 121-137.
57. Orata, Pedro T. "Measurement and Experimentation and Education for Independent Reconstruction," *Journal of Educational Research*, XXX (September 1936), pp. 1-13.
58. Price, Hazel Huston. *Securing Valid and Reliable Evidence of the Ability of the Adolescent Girl to Make Intelligent Decisions Concerning the Use of Personal Resources.* Columbus, Ohio: The Ohio State University (Ph.D. Thesis), 1938. 274 p. (Ms.)
59. Prieur, Marjorie. "Evaluation of Objectives in Family Relationships." *Educational Research Bulletin*, XVII (December 7, 1938), pp. 260-270.
60. Progressive Education Association, Evaluation in the Eight Year Study. Materials prepared by participants in the . . . Group of the Progressive Education Association Summer Workshop, Sarah Lawrence College, Bronxville, N. Y., July 2-August 13, 1937. Columbus, Ohio: Ohio State University, 1937. (Six mimeographed reports, in the fields of English, home economics, mathematics, science, social studies, and special groups.)
61. Pullias, E. V. "Study of Current Opinion Concerning Objective Tests," *Educational Method*, XVI (April 1937), pp. 348-356.
62. Ramseyer, L. L. "Factors Influencing Attitudes and Attitude Changes," *Educational Research Bulletin*, XVIII (January 4, 1939), pp. 9-14; 30.
63. Randall, J. A. "The Anecdotal Behavior Journal," *Progressive Education*, VIII (January 1936), pp. 21-26.
64. Raths, Louis E. "Appraising Certain Aspects of Student Achievement." *Guidance in Educational Institutions.* Thirty-Seventh Yearbook, Part I, National Society for the Study of Education (Bloomington, Illinois: Public School Publishing Company, 1938), Chapter 3, pp. 89-117.
65. Raths, Louis E. "Basis for Comprehensive Evaluation," *Educational Research Bulletin* (Ohio State University), XV (November 11, 1936), pp. 220-224.
66. Raths, Louis E. "Evaluating the Program of a School," *Educational Research Bulletin* (Ohio State University), XVII (March 16, 1938), pp. 57-84.
67. Raths, Louis E. "Techniques for Test Construction," *Educational Research Bulletin* (Ohio State University), XVII (April 13, 1938, pp. 85-114.
68. Raths, Louis E. "Experiencing the Realities of the Social Order, a Lincoln School Experiment: Some Evaluations of the Trip,"

Educational Research Bulletin, XVII (October 19, 1938), pp. 189-208.

69. Reene, Mary E. "If Not Testing—What?" *Educational Method*, XV (May 1936), pp. 403-406.
70. Remmers, H. H., editor. *Further Studies in Attitudes, Series II.* Studies in Higher Education, No. 31 (Lafayette, Indiana: Purdue University, 1936), 298 p.
71. Remmers, H. H., editor. *Further Studies in Attitudes, Series III.* Studies in Higher Education, No. 34 (Lafayette, Indiana: Purdue University, 1938), 151 p.
72. Remmers, H. H., editor. *Studies in Attitudes.* Studies in Higher Education, No. 26 (Lafayette, Indiana: Purdue University, 1934), 112 p.
73. Ross, F. J., and Heil, L. M. "Measuring the Listener's Attitude toward a Radio Art Appreciation Course," *Journal of Applied Psychology*, XXIII (February 1939), pp. 75-85.
74. Sayre, Jeanette. "A Comparison of Three Indices of Attitude toward Radio Advertising," *Journal of Applied Psychology*, XXIII (February 1939), pp. 23-33.
75. Scates, D. E., Sangren, P. V., Frutchey, F. P., Weidemann, C. C., Lindquist, E. F., and Courtis, S. A. "Educational Tests and Their Uses," *Review of Educational Research*, VIII (December 1938), pp. 495-596.
76. Scates, Douglas E. "The Essential Conditions of Measurement," *Psychometrika*, II (March 1937), pp. 27-34.
77. Shattuck, M. E. "In Defense of Appraisal: An Evaluating of Evaluation," *English Journal*, XXVIII (February 1939), pp. 95-102.
78. Spafford, Ivol. *Fundamentals in Teaching Home Economics.* (New York: John Wiley & Sons, 1935), Ch. 13.
79. Stalnaker, John M. "The Problem of the English Examination," *Educational Method*, XVII (October 1936), pp. 35-48.
80. Streitz, Ruth. "An Evaluation of Units of Work," *Childhood Education*, XV (February 1939), pp. 258-261.
81. "Summer Workshop in Secondary Education at Sarah Lawrence College, Bronxville, N. Y.," *Progressive Education*, XIV (April 1937), pp. 287-288.
82. Taba, Hilda. "Functions of Evaluation," *Childhood Education*, XV (February 1939), pp. 245-251.
83. Tyler, I. K. "Current Radio Research in Universities: Ohio State University," *Journal of Applied Psychology*, XXIII (February, 1939), pp. 195-199.
84. Tyler, Ralph W. "Appraising Progressive Schools," *Educational Method*, XV (May 1936), pp. 412-415.
85. Tyler, Ralph W. "Defining and Measuring Objectives of Progressive Education," *Educational Record*, XVII (January 1936), pp. 78-85. Same, condensed in *Educational Research Bulletin* (Ohio State University), XV (March 1936), pp. 67-72.

86. Tyler, Ralph W. "Evaluating the Outcomes of the Social Studies Curriculum," *The Social Studies Curriculum*, Fourteenth Yearbook (Washington: N. E. A. Department of Superintendence, 1936), pp. 312-343.
87. Tyler, Ralph W. "Objective Examination: Its Development and Application to Nursing Education," *American Journal of Nursing*, XXXVII (October 1938), pp. 1131-1138.
88. Tyler, Ralph W. "The Construction of Examinations in Botany and Zoology." *Service Studies in Higher Education* (Columbus, Ohio: Ohio State University Press, 1932), pp. 43-51. (Ohio State University Studies, Bureau of Educational Research Monograph No. 15).
89. Tyler, Ralph W. "Significance of a Comprehensive Testing Program." *Journal of Chemical Education*, XIV (April 1937), pp. 158-160.
90. Tyler, R. W. and others. "Coöperative general science test for college students," Coöperative Test Service. 1935.
91. Tyler, R. W. and others. "Coöperative General Achievement Test, Part II: A Survey Test in the Natural Sciences," for high school and college placement. Coöperative Test Service. 1937.
92. Tyler, R. W. "Evaluation: A Challenge to Progressive Education," *Progressive Education*, XII (December 1935), pp. 552-556.
93. Tyler, R. W. "Identification and Definition of the Objectives to Be Measured." Hawkes, H. E., et al., editors. *The Construction and Use of Achievement Examinations.* (Boston: Houghton Mifflin Company, 1936), pp. 3-16.
94. Tyler, R. W. "Needed Research in the Field of Tests and Examinations," *Educational Research Bulletin*, XV (September 1936), pp. 151-158.
95. Tyler, R. W. "Assumptions Involved in Test Construction," *Educational Research Bulletin*, XII (February 8, 1933), pp. 29-36.
96. Tyler, R. W. *Constructing Achievement Tests.* (Columbus, Ohio: The Ohio State University, 1934), 102 p.
97. Tyler, R. W. "Specific Techniques of Investigation: Examining and Testing Acquired Knowledge, Skill, and Ability," *Guidance in Educational Institutions.* 37th Yearbook, Pt. 2, National Society for the Study of Education (Bloomington, Illinois: Public School Publishing Company, 1938), pp. 341-355.
98. Tyler, R. W. "Techniques for Evaluating Behavior." In Pennsylvania University School of Education: *Educational Responsibilities of Today and Tomorrow* (Philadelphia: University of Pennsylvania, School of Education, 1935), pp. 348-357. (22d Annual Schoolmen's Week Proceedings.)
99. Tyler, R. W., and Frutchey, F. P. "Achievement tests in colleges and universities," *Review of Educational Research*, V (December 1935), pp. 491-501.

100. Van Antwerp, Harriet. "Teachers' Evaluation of the Effectiveness of Supervisory Activities," *Educational Method,* XV (May 1936), pp. 441-447.
101. Weedon, Vivian. "Research Needed in Interest Evaluation," *Educational Research Bulletin,* XVI (March 17, 1937), pp. 67-72, 84.
102. Wert, James E. "Twin Examination Assumptions," *Journal of Higher Education,* VIII (March 1937), pp. 136-140.
103. Wert, James E. *Educational Statistics.* (New York: McGraw-Hill Book Company, 1938), 305 p.
104. Wiseman, Edith C. "A Project for the Evaluation of School Broadcasts," *Educational Research Bulletin,* XVI (December 8, 1937), pp. 238-239, 244.
105. Wrightstone, J. W. "Test of Critical Thinking in the Social Studies in Grades 4, 5, and 6;" Forms A and B (Elementary School Series) with manual of directions and answer keys. Teachers College. 1939.
106. Wrightstone, J. W. "Coöperative Test of Social Studies Abilities." Experimental form 1936. High School. Coöperative Test Service.
107. Wrightstone, J. Wayne. "Tests for Evaluating Cardinal Objectives of Elementary Education," *Role of Research in Educational Progress.* Official Report of 1937 Meeting. (Washington, D. C.: American Educational Research Association, a department of the National Education Association, 1937), pp. 126-129.
108. Wrightstone, J. W. "Constructing an Observational Technique," *Teachers College Record,* XXXVII (October 1935), pp. 1-9.
109. Wrightstone, J. W. "Appraisal of Tests and Measures for Young Children," *Childhood Education,* XV pp. 252-257.
110. Wrightstone, J. Wayne. "New Tests for New Needs," *Educational Method,* XV (May 1936), pp. 407-411.
111. Wrightstone, J. W. "Achievement in Conventional and Progressive Public Schools." *Progressive Education,* XIII (May 1936), pp. 389-395.
112. Wrightstone, J. W., and others. "Measuring Social Performance Factors in Activity and Control Schools of New York City," *Teachers College Record,* XL (February 1939), pp. 423-432.
113. Wrightstone, J. Wayne. "Construction and Application of New Measures for the Social Studies." *Reconstructing Education Thru Research.* Official Report of 1936 Meeting. (Washington, D. C.: American Educational Research Association, a department of the National Education Association, 1936), pp. 164-169.
114. Wrightstone, J. W. *Appraisal of Experimental High School Practices.* (New York: Bureau of Publication, Teachers College, Columbia University, 1936), 194 p.
115. Wrightstone, J. W. *Appraisal of Newer Practices in Selected Public Schools.* (New York: Bureau of Publication, Teachers College, Columbia University, 1935), 117 p.

116. Wrightstone, J. Wayne. *Appraisal of Newer Elementary School Practices.* (New York: Teachers College, Columbia University, 1938), 221 p.

117. Wrightstone, J. Wayne. "Measuring Some Major Objectives of the Social Studies," *School Review,* XLIII (December 1935), pp. 771-779.

118. Wrightstone, J. Wayne. "Recent Trends in Social-Studies Tests," *Social Education,* I (April 1937), pp. 246-250.

119. Wrightstone, J. W. "Evaluation of School Broadcasts," in Institute for education by radio: *Education on the Air;* 9th Yearbook, 1938, pp. 87-92.

120. Wrightstone, J. W. "Generalized Procedure for Evaluating School Broadcasts." *Practical Values of Educational Research.* Official Report, 1938. (Washington, D. C.: American Educational Research Association, a department of the National Education Association, 1938), pp. 83-86.

121. Wrightstone, J. W. "Measuring Diverse Objectives and Achievement in Latin Teaching," *Classical Journal,* XXXIV (December 1938), pp. 155-165.

122. Wrightstone, J. W. "Appraisal of Newer Practices in Latin Teaching," *School and Society,* XLII (August 31, 1935), pp. 302-304.

123. Wrightstone, J. W. "Achievements in English in Activity Programs." *Elementary English Review,* XIII (March 1936), pp. 94-96.

124. Wrightstone, J. W. "An Appraisal of Newer Practices in Teaching French," *Modern Language Journal,* XIX (February 1935), pp. 369-374.

125. Wrightstone, J. W. "Evaluating Practices in Health and Physical Education," *American Physical Education Association Research Quarterly,* XVII (March 1936), pp. 26-35.

126. Wrightstone, J. W. "How Can the Thirty-two States Evaluate Curriculum Revisions?" *Clearing House,* XI (February 1937), pp. 333-337.

127. Wrightstone, J. Wayne. "Measuring the Attainment of Newer Educational Objectives," *Appraising the Elementary-School Program.* Sixteenth Yearbook. (Washington, D. C.: Department of Elementary School Principals, National Education Association, 1937), pp. 493-501.

128. Wrightstone, J. W. "Measuring Intangibles in Progressive Schools," *Progressive Education,* XII (February 1935), pp. 95-97.

129. Zahn, D. Willard. "The Anecdotal Record in Relation to Character Development," *Education for Dynamic Citizenship.* Twenty-Fourth Annual Schoolmen's Week Proceedings. (Philadelphia: School of Education, University of Pennsylvania, 1937), pp. 294-300.

A STUDY OF THE EXTENT OF MEASUREMENT OF DIFFERENTIAL OBJECTIVES OF INSTRUCTION

T. R. McConnell*
University of Minnesota

Editor's note: It has been frequently said in recent years that measurement should be with reference to the validated purpose of education. Accepting this emphasis as a point of departure the author presents important data on the relation of measured information and measured application.

The first step in educational measurement and appraisal, it is now generally agreed, is to draw up a comprehensive list of educational objectives. The next procedure is to state these outcomes in terms of behavior. Techniques of several sorts may then be devised to provide estimates of the extent to which students have developed behavior patterns relevant to the objectives.

This attack upon the problems of measurement has emphasized the variety of outcomes of educational experiences and produced new instruments of appraisal. One of the first results of the movement was to supplement tests composed almost exclusively of items of specific information with measures of the ability to relate facts, to apply facts and principles to new situations, and to draw inferences from data. Tyler's significant study of the relationships between the results of tests of recall, application, and inference in a variety of subject matter fields not only stimulated attempts to measure more than informational outcomes, but also called attention to the necessity for cultivating the higher mental processes by appropriate learning activities.[1] He found that a "large number of students studying a variety of subjects did not develop corresponding degrees of facility in mere recall and facility in the higher mental processes of applying principles and drawing inferences." In 16 examinations from 11 different subjects, he found that the correlations between tests requiring recall of information and those requiring both recall and application of principles, ranged, when corrected for attenuation, from .31 to .58. Most of the coefficients hovered about .45.

At the University of Minnesota, the University Committee on Educational Research and the General College coöperated in the development

* With the assistance of Marian Wilder and N. Fattu, Statisticians for the University Committee on Educational Research; paper presented at the St. Louis meeting of the American Educational Research Association, 1940.

[1] Tyler, R. W. "The Relation Between Recall and Higher Mental Processes" in Judd, C. H., *Education as Cultivation of the Higher Mental Processes* (New York: Macmillan, 1936), pp. 6-17.

of course and comprehensive examinations between 1933 and 1938. Although other objectives were also measured, the examinations stressed the following outcomes, since they were the objectives most frequently mentioned by instructors in certain General College areas: knowledge of vocabulary, knowledge of facts and principles, and the application of facts and principles in new situations.[2]

In an earlier study on the differential functions of examinations, Johnson reported higher correlations between knowledge of facts and principles and the ability to apply them to new situations than Tyler had obtained. The coefficient (corrected for attenuation) between two such sections of a comprehensive examination in Human Biology was .77. In a comprehensive examination on Physics and Chemistry, the correlation was .87, and in a comprehensive in Basic Wealth, it was .89.[3]

It seemed desirable to conduct further studies of certain comprehensive examinations prepared for the General College under the auspices of the University Committee in order to reveal the extent to which these tests had measured differential outcomes. This report summarizes the results for the Physical Science Comprehensive given to 73 General College students in the spring of 1938. This examination was selected for the investigation because it was considered one of the best which had been prepared under the Committee's auspices, and because its subdivisions, both in subject matter and functional objectives, seemed the most clear-cut.

The examination was divided into three major subject matter divisions: (A) Physics, (B) Acoustics and Astronomy, and (C) Chemistry. Each of the three subject matter divisions included items designed to measure (1) knowledge of vocabulary and (2) knowledge of facts and principles; under the Physics and Chemistry sections, there were also (3) items which purported to test students' ability to apply facts and principles.

Table I gives the total number of items in each of the eight parts of the test and the examination as a whole; the mean score, standard deviation, and the reliability coefficients for each section. In this and the sub-

[2] Eurich, A. C., and Johnson, P. O. "The Experimental Examination Program in the General College," in University Committee on Educational Research, *The Effective General College Curriculum as Revealed by Examinations.* (Minneapolis: The University of Minnesota Press, 1937), pp. 31-42.

[3] Johnson, P. O. "Differential Functions of Examinations," in University Committee on Educational Research, *Studies in College Examinations,* (Minneapolis: University Committee on Educational Research, University of Minnesota, 1934), pp. 43-50.

sequent tables, the capital letters (A, B, C) refer to the subject matter sections of the test, and the figures (1, 2, 3) to the outcomes measured. Thus A1 refers to the vocabulary subdivision of the Physics section.

TABLE I

MEAN SCORES, STANDARD DEVIATIONS, AND RELIABILITIES OF THE EIGHT SECTIONS OF THE PHYSICAL SCIENCE COMPREHENSIVE

(N = 73)

	Sections of the Examination								
	A1*	A2	A3	B1	B2	C1	C2	C3	Total Test
Number of Items	65	75	40	15	65	60	95	44	459
Mean Score	38.89	34.47	21.30	6.30	29.38	37.66	41.66	14.32	223.97
Standard Deviation	9.82	8.79	5.11	2.43	6.66	9.29	11.58	4.98	47.43
Reliability (r_{11})	.87	.87	.74	.65	.77	.90	.90	.76	.97

* The capital letters refer to the subject matter sections, and the figures to the outcomes measured. Thus A1 means Physics-vocabulary.

The extent of differential measurement was first studied by the method of item discrimination. The assumption underlying this procedure was that an item designed to measure factual information should discriminate among students of high and low achievement better when the criterion was total score on its own division of the examination than when the criterion was the total score on a division devised to test some other objective, such as the ability to apply facts and principles to new situations.

Therefore, the discriminating power of each item of the examination was computed, using the total score on each of the eight parts and the entire test in turn as the criterion for dividing the 73 students into high, middle, and low groups in achievement. Approximately one-third of the total number of students was included in each of these groups, although the division was made each time so as to include all cases with the same score. This caused the number of students in the three groups to change slightly from one section to another. Table II gives the number of cases in the high, middle, and low groups for each section of the examination, and for the test as a whole.

The discriminating power of a test item was represented by the ratio of the difference in percentages of students in the high and low groups who answered the item correctly, to the standard error of that difference.[4]

[4] Elveback, M. L. "Conditions Affecting the Differentiating Capacity of Test Items," in University Committee on Educational Research, *The Effective General College Curriculum as Revealed by Examinations*, pp. 325-332.

TABLE II

NUMBER OF CASES IN HIGH, MIDDLE, AND LOW GROUPS FOR ITEM ANALYSES
(N = 73)

	Sections of the Examination								Total Test
	A1	A2	A3	B1	B2	C1	C2	C3	
Number in High Group	21	25	25	16	26	25	27	28	25
Number in Middle Group	23	24	25	39	22	25	24	22	23
Number in Low Group	29	24	23	18	25	23	22	23	25

$$\text{relative deviate} = d = \frac{P_1 - P_2}{\sqrt{\frac{P_1(100 - P_1)}{N_1} + \frac{P_2(100 - P_2)}{N_2}}}$$

An item was called discriminating if the relative deviate was as large as two. This, of course, is an arbitrary standard, and is perhaps rather severe.[5] It conforms, however, to the practice which has been used by the University Committee on Educational Research.

Relative deviates were computed for each item, using the total score on its own and each of the other seven sections and on the entire test in turn, as criteria for separation into high and low achievement groups. Space limitations preclude the presentation of the tables of these values. They have been placed in the files of the University Committee on Educational Research, however, and will be made available to interested persons.

Table III summarizes the number of items in each section which were discriminating (according to the standard noted above) with respect to the total score on each of the parts and on the entire test. Of the 65 items in section A1 (Physics-vocabulary), for example, 47 were discriminating with respect to the total score on their own part; 39 of them were discriminating on the Physics-information (A2) criterion; 27, on the Physics-application criterion; 19, on the Acoustics-Astronomy-vocabulary criterion; 29, on the Acoustics-Astronomy-information criterion; 20, on Chemistry-vocabulary; 27, on Chemistry-information; 13, on Chemistry-application; and 41, on the criterion of the total test. When a test is relatively short, the differentiating power of an item may be affected by the inclusion of the score on that item in the criterion score. The last column of Table III shows the number of items which were differentiating in their own sections when the score on an item under

[5] If the standard is reduced to a relative deviate of one, the results do not change the general findings of this paper.

consideration was excluded from the total score on its section. The use of this correction in the case of the Physics-vocabulary part reduced the number of differentiating items in that section from 47 to 43.

An inspection of Table III reveals that when no correction was made for the inclusion of the score on an item in the total score of its section, the largest number of differentiating items was obtained in all sections by using as the criterion the total score on the section of which the item was a part. However, when the score on an item was excluded from the total score on its own section, this criterion gave the largest number of differentiating items in three of the eight sections. In two other sections, total score on the test gave the largest number; in two sections, some other part of the test yielded the largest number, and in one case the numbers given by the items' own section score and score on some other part were equal. The second largest number of differentiating items was given in four cases by the total score on the test; in three cases by the score on some other section of the test; and in the remaining case there was a tie between the items' own section score and that on another part.

TABLE III

NUMBER OF ITEMS IN EACH SECTION WHICH WERE DISCRIMINATING ON EACH CRITERION SCORE

Test Section	Total Number of Items	Number of Differentiating Items for Each Criterion Section and for Entire Test									Number Discriminating Items with Corrected Criterion Score*
		A1	A2	A3	B1	B2	C1	C2	C3	Total	
A1	65	47	39	27	19	29	20	27	13	41	43
A2	75	27	42	32	9	19	14	16	22	34	37
A3	40	11	17	24	7	15	6	6	10	14	17
B1	15	5	2	3	9	5	2	2	2	3	6
B2	65	17	20	19	13	28	8	8	13	16	18
C1	60	21	19	17	9	10	45	37	25	32	30
C2	95	22	25	21	12	17	37	52	36	43	30
C3	44	9	11	9	4	11	16	11	23	19	10

* i.e., when score on an item under consideration was excluded from total score on its own section.

The differences between the discriminating power of items when the criterion was total score on the part to which they belonged, and when the criterion was the total score on each other part of the test, were computed for all sections of the examination, and were expressed as differ-

ences between the various relative deviates. These data, not reproduced here, are on file in the office of the University Committee on Educational Research. Table IV, however, gives a summary of the basic distributions. It shows the mean differences of the discriminating powers (in terms of relative deviates) of the items of each section when each of the other seven parts was used in turn as the criterion. It also gives the standard deviations of these differences, as well as "t" and the related probability value. The score on an item was always excluded from the appropriate criterion total score.

An inspection of Table IV reveals a tendency for smaller differences in differentiating powers between subdivisions of the same subject matter than between corresponding subdivisions of different subject matter. For example, the mean differences between A1 and A2, and A1 and A3, are .26 and .73 respectively, while the mean differences between A1 and

TABLE IV

SIGNIFICANCES OF MEAN DIFFERENCES IN ITEM DISCRIMINATION FROM ONE CRITERION TO ANOTHER

Test Section	N		Criterion Sections: A1	A2	A3	B1	B2	C1	C2	C3
A1	65	Mean		.26	.73	1.31	.75	1.08	.91	1.29
		S. D.		1.39	1.26	1.40	1.18	1.34	1.25	1.33
		t		1.52	4.62	7.49	5.10	6.42	5.83	7.74
		P		.13	.01	.01	.01	.01	.01	.01
A2	75	Mean	.18		.10	.95	.35	.79	.53	.51
		S. D.	1.03		.98	1.48	1.17	1.42	1.22	1.15
		t	1.46		.90	5.51	2.55	4.79	3.73	3.83
		P	.14		.37	.01	.01	.01	.01	.01
A3	40	Mean	.61	.32		1.44	.50	1.11	1.27	.94
		S. D.	1.35	1.22		1.71	1.30	1.58	1.59	1.38
		t	2.84	1.65		5.26	2.40	4.42	4.97	4.26
		P	.01	.10		.01	.02	.01	.01	.01
B1	15	Mean	1.04	1.19	1.64		.92	1.11	1.53	1.39
		S. D.	2.11	1.89	2.35		2.10	1.62	1.51	1.73
		t	1.85	2.34	2.62		1.64	2.55	3.80	3.00
		P	.08	.04	.02		.13	.03	.01	.01
B2	65	Mean	.16	— .001	— .02	.26		.52	.50	.29
		S. D.	1.14	1.15	.98	1.41		1.49	1.48	1.43
		t	1.09	— .01	— .20	1.45		2.80	2.71	1.63
		P	.27	.99	.84	.15		.01	.01	.10
C1	60	Mean	.52	.50	.64	.81	.90		— .26	.14
		S. D.	1.43	1.20	1.37	1.38	1.45		1.02	1.16
		t	2.77	3.22	3.60	4.48	4.80		—1.93	.95
		P	.01	.01	.01	.01	.01		.05	.34
C2	95	Mean	.32	.11	.33	.89	.47	— .09		— .19
		S. D.	1.31	1.35	1.65	1.68	1.74	1.02		1.10
		t	2.37	.80	1.93	5.12	2.62	— .84		—1.66
		P	.02	.42	.05	.01	.01	.40		.10
C3	44	Mean	— .09	— .28	— .25	.41	— .21	— .55	— .51	
		S. D.	.95	1.00	.94	1.23	.96	1.06	.96	
		t	— .59	—1.80	—1.78	2.19	—1.42	—3.37	—3.51	
		P	.55	.07	.08	.03	.16	.01	.01	

B1, and A1 and C1, are 1.31 and 1.08, respectively. There is a corresponding tendency for the "t" to be insignificant between the subdivisions (i.e., vocabulary, facts, and applications) of the same subject matter, and for "t" to be significant when the differences are between corresponding outcomes (vocabulary, for example) of different subject matter. Thus the differences between the outcomes within the same subject seem to be smaller than those between the subjects within the same outcome. That is, items have more nearly the same differentiating power when the criterion used is either vocabulary, information, or application sub-sections of the same subject (say, Physics) than when the criterion is the Physics, Astronomy, or Chemistry sections of the same outcome (say, vocabulary).

The same tendencies can be observed from a study of the intercorrelations between the eight sections of the examination. Table V shows these relationships; the correlations given above in each case are the uncorrected values, while those below have been corrected for attenuation. One immediately wonders whether the average of all the correlations within the subject matter areas (such as $r_{A1.A2}$, $r_{A1.A3}$, $r_{A2.A3}$, $r_{B1.B2}$, etc.,) is higher than the average of all the correlations within the same type of outcome (such as A1.B1, A2.B2, etc.). To answer this question, each of the correlations was transformed to Fisher's z, and combined (as outlined above) and averaged. The average of all the correlations within

TABLE V

INTERCORRELATIONS BETWEEN SECTIONS, UNCORRECTED AND CORRECTED FOR ATTENUATION

($N = 73$)

	A2	A3	B1	B2	C1	C2	C3
A1	.76*	.60	.38	.58	.53	.58	.51
	.87	.75	.51	.71	.60	.65	.62
A2		.74	.35	.63	.58	.65	.63
		.92	.47	.77	.66	.74	.77
A3			.19	.67	.51	.53	.59
			.27	.89	.63	.65	.79
B1				.43	.42	.29	.32
				.61	.54	.38	.45
B2					.48	.49	.59
					.58	.59	.78
C1						.80	.71
						.89	.86
C2							.80
							.96

* The numbers above in each are the uncorrected values; those below, corrected for attenuation.

the three subject matter areas was .87. The average of all the correlations within the three outcome divisions, vocabulary, information, and application, was .66. The difference between these two average correlations had a "t" of 3.15, which is significant at the one percent level. One concludes, therefore, that for this examination, there was greater homogeneity within subject-matter sections than within groupings based on outcomes.

A comparable analysis of a carefully constructed and classified test in Human Development and Personal Adjustment yielded substantially the same results as those reported in this paper.[6] The intercorrelations and item discrimination analyses in both tests have shown much less evidence of differential measurement than Tyler's earlier study. What might account for the inconsistency?

It has been suggested that a consistent instructional emphasis on the application of facts and principles will lead to a closer relationship between measured knowledge and its application than obtained when teaching and testing stressed mainly the acquisition of specific and relatively discrete information. It is true that the General College has attempted from the beginning to make its teaching functional. We have no extensive data with which to compare the results of tests conducted at the beginning and end of the six years of the General College examination project. There is a bit of evidence in the earlier study by Johnson referred to above. His intercorrelations were not greatly different from those in this study.

Research on the transfer of training suggests that the relationship between information and application would change with the difficulty of the application problems, and also with the degree of similarity between the original learning situations and those to which the facts and principles were to be applied in the examination. One might expect, also, to find a lower correlation between a knowledge of facts and principles tested verbally and applications tested in performance, than between knowledge and application when both were measured in purely verbal situations.

The adequacy with which knowledge of facts and principles was measured would also probably affect the relationships very materially. Lindquist has pointed out that objective examinations have often called merely for detailed and discrete items of information, and that answering the

[6] This study was begun by Dr. J. E. Anderson and completed by the University Committee on Educational Research.

items correctly often required only "rote learning rather than real understanding of this information." Under such circumstances, one might expect to find a low relationship between acquisition of information and the ability to make applications and inferences. Lindquist has also contended that "the subject matter of any course of instruction . . . may be considered as consisting of a body of purely descriptive or encyclopedic facts, of relationships between these facts, and of ideas, generalizations, principles, and methods of procedure based on facts and relationships. All of this may, in a sense, be thought of as 'information'."[1] It is clear that if one measures information in terms of real understanding rather than in terms of rote memory, the test of knowledge will approach, at least, the character of a test of application and inference.

If such considerations as those just discussed are of any real significance in determining the relationship between measured knowledge and measured application of that information, they would suggest that it might be possible to construct tests which would show either a high or low correlation between the two outcomes depending on the nature of the material included, and the manner of its organization into test items.

There is, of course, nothing in the data or the hypotheses presented in this paper that would disparage the attempt to measure application and inference. Any emphasis which can be given in curriculum and instruction (and the nature of examinations probably will influence both) on the relationships of important facts and their use in solving significant problems is crucial in the educative process.

[1] Hawkes, H. E., Lindquist, E. F., and Mann, C. R. *The Construction and Use of Achievement Examinations.* (Boston: Houghton Mifflin, 1936), pp. 81-83.

THE WISCONSIN STUDY OF TEACHING ABILITY

A. S. Barr
University of Wisconsin

Editor's note: As is the teacher, so is the school. The author presents a preliminary report upon a study of teaching efficiency made at the University of Wisconsin.

PURPOSE OF THE INVESTIGATION

Many persons engaged in the institutional and in-service training of teachers have come to realize that the selection, guidance, training, and evaluation of teachers must be put upon a more substantial basis. The purpose of this investigation is to throw light, if possible, upon the qualities essential to success in teaching, the means of identifying these qualities in specific persons and situations, and the changes that may take place in them with training. More specifically the study attempts to answer the following questions: (1) What are the qualities essential to success in teaching, particularly for teachers of the social studies in the 7th and 8th grades of rural schools in Wisconsin? (2) How valid and reliable are certain instruments commonly employed in the measurement of teaching ability? And (3) what changes may be made in the qualities ordinarily considered essential to success in teaching through training, particularly through the training of teachers in service? Besides these more specific purposes, it is hoped that this investigation may throw light upon the nature and organization of human abilities in general as exemplified in teaching. There are many other important problems in this field but those here enumerated seemed most appropriate to the conditions under which the investigation was conducted.

PLANNING THE INVESTIGATION

One hears from time to time a desire expressed that persons with similar interests might have more opportunities to work together coöperatively in the solution of the complex problems of education. The plan of research and the research here reported is the product of the mental efforts of many persons working together in a coöperative adventure, with representatives from the department of public instruction of the State of Wisconsin, the University of Wisconsin Department of Education, and many teachers, principals, supervisors and superintendents. During the school year of 1934-35 a group of some twenty

persons from the above named groups met every other Saturday morning throughout the year to discuss possible approaches to the solution of the problem. The result of this year's work was a tentative statement of the problem and a procedure for further investigation. This tentative formulation was next turned over during the academic year of 1935-36 to a graduate seminar of some twenty-six students for critical analysis and revision. During the first semester of this year the procedure was rewritten in light of the critical comments of the group and previous investigations; during the second semester the procedure was tried out in a preliminary investigation. The procedure of the investigation here reported was formulated at the end of the second year of planning and as a result of two year's effort in this direction. I should like to emphasize again the coöperative character of this investigation. The investigation was and is now open to any and all persons who may desire to participate with full recognition for all.

THE IMPORTANCE OF THE INVESTIGATION

Aside from the fact that a better insight into the nature and organization of human abilities, in general, may be had from a study of these abilities in some particular field such as teaching, this investigation aims primarily to be of value to the teaching profession. There are approximately one million and a quarter teachers in this country who teach thirty million pupils. The information sought herein should be of value, first, in the guidance of persons into or away from the profession of teaching by making known the qualities essential to success in teaching. The teacher brings her whole self to teaching and it is apparent that not all persons are equally well fitted for teaching and to choose those that have special promise of succeeding as teachers involves a knowledge of the abilities, knowledges, skills, attitudes, ideals, interests, appreciations, traits and forms of behavior associated with teaching efficiency. In the second place, the information sought herein should be of value in the institutional and the in-service training of teachers by indicating the kinds of training needed to produce better teachers: the constituents of effective teaching and how these constituents may be acquired. Finally, the information sought herein should be of value in the further refinement of the instruments for the measurement of teaching efficiency. These instruments furnish the means by which teachers and the content, methods, and materials employed in their education may be evaluated. To secure better teachers, then, one must, first, admit into teacher train-

ing institutions only those individuals who show capacity for teaching; secondly, one must furnish these individuals with the kinds of training and guidance that will provide them with the essential constituents for effective teaching; thirdly, one must base the methods of training teachers upon information about how teachers learn to teach and the conditions for effectively learning; and lastly one must be everlastingly alert to the products of teacher training institutions and of programs designed for the in-service training of teachers, to make certain that they live up to their expectations. To assure oneself that they do live up to expectations one must measure repeatedly the outcomes of teacher education. The investigation herein outlined would seem to furnish at least one attack upon a fundamental problem of education.

THE GENERAL PLAN OF THE INVESTIGATION

It appeared that the purposes of this investigation might best be served by the use of the controlled case study statistical technique of research applied under normal classroom conditions. The application of this technique involves, in the collection of data, the following steps: (1) the definition of a task to be performed in terms of the goals to be attained; (2) the determination of the changes produced in pupils by the application of appropriate instruments of measurement before and after instruction; (3) the determination of the specific respects in which the outcomes of instruction are satisfactory and unsatisfactory; (4) the determination through the use of appropriate diagnostic techniques of the probable causes of satisfactory and unsatisfactory pupil growth and achievement; (5) the formulation of an improvement program designed to remove the supposed cause or causes of unsatisfactory pupil growth and achievement; and (6) the evaluation of the performance of the pupils to ascertain whether the performance has or has not been improved.

The investigation here reported was carried out with seventh and eighth grade teachers of citizenship in non-departmentalized one room rural and state graded schools in the state of Wisconsin. The principal criterion of teaching efficiency employed in this investigation was a composite of measures of pupil growth and achievement. A composite of the scores on three rating scales twice applied by three observers and a composite of the scores on eleven tests of qualities commonly associated with teaching success constituted secondary criteria. The development of these criteria involved first, the taking of two samples, each three weeks

in length, of each teacher's teaching at six month intervals, one sample near the beginning and one near the end of the school year (tests were given before and after the teaching of each sample unit); and, secondly, the application of certain overall tests to the pupils at the beginning and end of the school year. In taking the samples of teaching, each teacher was asked to perform certain standard tasks within the regular course in citizenship defined in terms of the desired outcomes of instruction, each task being of equal difficulty, involving the instruction of pupils of equal capacity under comparable conditions. The overall tests related to the more general purposes of the course in citizenship, the teachers of which were here under investigation. It was thought that these two types of information might supplement each other. The list of the tests given to the pupils is given below:

Measures Applied to the Pupils

1. Wilbur Brookover, *A Test of Teacher-Pupil Relationship.* Forthcoming issue, *Journal of Educational Research.*
2. Howard C. Hill, *A Test in Civic Information.* Bloomington, Illinois: Public School Publishing Co., 1926.
3. Howard C. Hill, *A Test in Civic Attitude.* Bloomington, Illinois: Public School Publishing Co., 1926.
4. Howard C. Hill, and Howard E. Wilson, *A Test in Civic Action.* Bloomington, Illinois: Public School Publishing Co.
5. F. Kuhlmann, and Rose Anderson, *Kuhlmann-Anderson Tests.* Minneapolis, Minn., and Philadelphia, Pa., Education Test Bureau, Inc., Fourth Edition, Grades VII-VIII, 1927.
6. Verner Martin Sims, *Sims' Score Card for Socio-Economic Status,* Form C. Bloomington, Illinois: Public School Publishing Co., 1927.
7. Arthur E. Traxler, *Traxler Silent Reading Test,* Form 1. Bloomington, Illinois: Public School Publishing Co., 1934.
8. *Unit Test on Community Planning* (Informal objective test developed by the research staff), Form B.
9. *Unit Test on Health* (Informal objective test developed by the research staff), Form B.
10. J. Wayne Wrightstone, *Scale of Civic Beliefs,* Form A. Yonkers: World Book Co., 1935.
11. J. Wayne Wrightstone, *Abilities to Organize Research Materials,* Revised Form. Yonkers: World Book Co., 1935.
12. J. Wayne Wrightstone, *Applying Generalizations to Social Studies Events.* Yonkers: World Book Co., 1935.

Certain of these tests were tests of achievement and others of abilities considered important principally in the equation of pupils. Besides these

tests given to the pupils were a number of tests and ratings applied to the teachers. A list of these measures is given below:

Measures Applied to the Teachers

1. H. C. Almy, and Herbert Sorenson, *Almy-Sorenson Rating Scale for Teachers.* Bloomington, Illinois: Public School Publishing Co., 1930.
2. A. S. Barr, *Teacher's Performance Record.* Bloomington, Illinois: Public School Publishing Co., 1940.
3. A. S. Barr, and Robert H. Steiner, *A Test of Professional Judgment.* Bloomington, Illinois: Public School Publishing Co., 1940.
4. Robert C. Bernrueter, *The Personality Inventory.* Stanford University, California: Stanford University Press.
5. George W. Hartman, *Social Attitudes of Secondary-School Teachers.* Materials prepared for the John Dewey Society for the Study of Education and used by the special permission of George W. Hartman, Columbia University.
6. Vergil E. Jackson, *A Test of Social Proficiency.* Forthcoming issue of the *Journal of Experimental Education.*
7. Milton B. Jensen, *Stanford Educational Aptitude Test.* Stanford University, California: Stanford University Press, 1928.
8. Robert D. Leigh, Peter H. Odegard, and Ben D. Wood, *American Council Civics and Government Test,* Form B for High Schools and Colleges. Yonkers: World Book Co., 1929.
9. Alfred S. Lewerenz, and Harry C. Steinmetz, *Orientation Test, Concerning Fundamental Aims of Education, for Use in High School and College,* 1935 Revision. Los Angeles, California: Southern California Book Depository, 3636 Beverly Boulevard, 1935.
10. C. F. Malmberg and others, *Psychological Examination, Prepared for Teachers College Personnel Association,* Form C, 1938 edition. Normal, Illinois: Illinois State Normal University, 1938.
11. *Michigan Education Association Teachers Rating Card,* Long form, Human Scale Method. Lansing, Michigan: Michigan Education Association.
12. Elizabeth Morris, *Morris Trait Index L.* Bloomington, Illinois: Public School Publishing Co., 1936.
13. *Personality Rating Scale,* unpublished materials, Department of Education, University of Wisconsin, 1936.
14. Mabel Rudisill, *A Scale for Measuring the Personality of Elementary School Teachers,* unpublished materials, Department of Education, University of Wisconsin, 1931.
15. Verner Martin Sims, *Sims Score Card for Socio-Economic Status,* Form C. Bloomington, Illinois: Public School Publishing Co., 1927.
16. L. L. Thurstone, and Thelma Gwinn Thurstone, *American Council on Education Psychological Examination for College Freshmen,* 1936 Edition. Washington: American Council on Education, 1936.

17. T. L. Torgerson, *Torgerson Diagnostic Teacher Rating Scale of Instruction Activities.* Bloomington, Illinois: Public School Publishing Co., 1930.
18. T. L. Torgerson, Bjarne Ullsvick, and Laurence Wahlstrom, *A Test of Teacher-Pupil Relationship.* Unpublished materials, University of Wisconsin.
19. J. N. Washburne, *The Social Adjustment Inventory.* Syracuse, New York: Syracuse University, 1936.
20. Tressa C. Yeager, *Scale for Measuring Attitude Toward Teachers and Teaching Profession.* Reproduced, by special permission of the author, from *An Analysis of Certain Traits of Selected High School Seniors Interested in Teaching,* Contributions to Education, No. 660, Bureau of Publications, Teachers College, Columbia University, 1935.

Other information collected with reference to each teacher, included sound records, two for each of twenty-five teachers; samples of pupil's work; and teaching outlines. All teachers were given the unit pretests along with the pupils, and each teacher filled out a detailed information blank. Inasmuch as this study is a coöperative investigation, some variation will be found in the procedures and measures employed in the various individual investigations.

SPECIAL INVESTIGATIONS INCLUDED HEREIN

As has been said the investigation here reported represents the efforts of many persons. Special investigations have been made as follows:

1. A study of the validity and reliability of certain instruments commonly employed in the measurement of teaching ability (first investigation). Leon C. Rostker.
2. A study of the qualities commonly associated with teaching success (first investigation). L. H. Mathews.
3. A study of the validity and reliability of certain instruments commonly employed in the measurement of teaching ability (second investigation). R. F. Rolfe.
4. A study of the validity and reliability of certain new type instruments commonly employed in the measurement of teaching ability (third investigation). C. V. LaDuke.
5. A study of the validity of certain instruments commonly employed in the measurement of teaching efficiency using composite criteria. Algot Smith.
6. A study of the relationship between performance in the classroom and teaching efficiency. Albert E. Harris.
7. Case studies of teaching efficiency. Lucile Madsen.
8. Characteristic differences in the teaching performance of good and poor teachers. G. E. Carlson.

9. The social impact of teachers upon pupils. Russell Bollinger.
10. A study of the relationship between specific teacher activities and teacher efficiency. C. D. Jayne.
11. A study of the improvability of teachers in service under supervision. Clarence E. Von Eschen.
12. A study of the speech factor in teaching efficiency. William B. McCoard.
13. A study of teacher-pupil relationship as a factor in teaching efficiency. Wilbur Brookover.
14. A study of the teacher's personality as a factor in teaching success. Raymond E. Gotham.
15. A factor analysis of teaching ability. Alvin G. Hellfritzsch.
16. Special problems in the measurement of pupil growth and achievement. Naomi Livesay and Gerald Vaughan.
17. A study of the relationship between levels of attainment for different measures of teaching ability and teaching efficiency. Maurice A. Hannon.
18. Social experiences, training, and teaching efficiency. Robert Steiner.
19. An annotated bibliography on the measurement of teaching efficiency. Thomas L. Faucett.

THE TEACHERS AND PUPILS STUDIED

The main body of data for the investigation here reported consist of measures of six groups of teachers and pupils (two of the six groups of teachers were followed through a two year period).[1] The groups were as follows: (1) a group of 24 teachers teaching in state graded schools with 223 pupils; (2) a group of 21 teachers with 197 pupils chosen from the above for a special one-year follow-up study; (3) a group of 47 teachers teaching in one room rural schools with 385 pupils; (4) a group of 24 teachers with 194 pupils chosen from the above for a special study of the improvability of teachers in service under supervision; (5) a group of 34 teachers with 200 pupils employed in a check investigation; and (6) a group of 17 teachers and 400 pupils in three communities employed in a special study of the social impact of the teacher upon the pupil. All in all, some 167 teachers and 1,599 pupils in one room and state graded village schools in the State of Wisconsin were employed in the investigation. More detailed descriptions of the teachers and pupils studied will be given in the special reports to follow.

DEFINING THE TASK (FOR TEACHERS AND PUPILS)

The task to be performed by the teachers and pupils participating in the investigation was defined in terms of certain course and unit

[1] Certain of the studies employed supplementary data from other than rural schools.

objectives placed in the hands of the teachers. These directions as given to the teachers are reproduced below:

The teacher should aim in the citizenship course to develop the pupil's ability to discharge his socio-civic responsibilities with intelligence and efficiency. This will involve: (a) *efficiency in human relationships:* the ability to get along with other people; the ability to work coöperatively with others in group enterprize; the ability to subordinate one's individual gain to group welfare where one conflicts with the other, etc.; (b) *understanding of the fundamentals of socio-civic relationships:* knowledge of one's civic rights, duties, and responsibilities; knowledge of social, economic, and political principles and practices; knowledge of moral, ethical, and religious conventions; understanding of the principles and practices of the government under which one lives, etc.; (c) *the mastery of the tools for effective social behavior:* the ability to see and solve problems; the ability to collect relevant data, make the necessary analyses, and reach judgments based upon fact; the ability to suspend judgment, maintain an open mind, and view solutions critically in arriving at decisions, etc.; and (d) *interest in socio-civic relationships, activities, and responsibilities:* willingness to accept socio-civic responsibilities, the faithful discharge of these responsibilities; and an attitude of tolerance toward the opinions and actions of others.

UNIT ASSIGNMENT NO. 1: SAFEGUARDING PUBLIC HEALTH

The unit on safeguarding public health is to include a treatment of the following topics: (1) securing pure air, water, food and sunshine; (2) disposing of wastes; (3) providing desirable housing facilities; (4) caring for the physically and mentally sick; and (5) recreational opportunities. The assignment is made not in terms of subject matter to be taught, but in terms of the goals of instruction to be sought. With this thought in mind, the teacher's instruction should be directed toward the following objectives: (1) To acquire the kinds and amounts of information essential to the understanding of the problems and issues involved in safeguarding public health. (2) To develop skill in forming judgments about this subject: in the recognition of important problems and issues; in the analysis, classification and systematization of information; and in reaching reasoned judgments supported by fact. (3) To develop desirable attitudes relative to safeguarding public health; to bring pupils to see the importance of the problem; to interest pupils in the means of safeguarding public health; and to develop the necessity for coöperative action. (4) To lead the pupils, individually and coöperatively, to some positive action relative to safeguarding public health in their immediate environment in keeping with their abilities.

The teacher is free to employ whatever means she deems appropriate for the attainment of these objectives, that is, she is free to use any subject matter, methods, or materials that she may think best in the attainment of the above objectives.

UNIT ASSIGNMENT NO. 2: COMMUNITY PLANNING

The unit on community planning is to include a treatment of the following topics: (1) the layout of streets; (2) building zones; (3) beautifying the community; and (4) keeping the community clean; and (5) recreational facilities. The assignment is made not in terms of subject matter to be taught, but in terms of the goals of instruction. With this thought in mind, the teacher's instruction should be directed toward the following objectives: (1) To acquire the kinds and amounts of information essential to the understanding of the problems and issues involved in planning the community. (2) To develop skill in forming judgments about this subject: in the recognition of important problems and issues; in the collection of information relating to these problems and issues; in the analysis, classification and systematization of information; and in reaching reasoned judgments supported by fact. (3) To develop desirable attitudes relative to planning the community: to bring pupils to recognize the importance of the problem; to interest pupils in the means of planning the community; and to develop the necessity for coöperative action. (4) To lead the pupils, individually and coöperatively, to some positive action in their immediate environment in keeping with their abilities, relative to carrying into effect plans discussed.

An attempt was made to validate the purposes of the course in citizenship studied in this investigation through an examination of current statements[2] of the purposes of education and two special investigations[3] of the purposes of instruction in citizenship. The form in which these objectives were stated was that most generally employed in the literature, but not, however, without some misgivings. There appear to be two quite different forms in which the purposes of education may be stated, namely, (a) statements in terms of pupil behavior; and (b) statements in terms of controls over behavior. These latter are of two sorts: (a) desired traits, qualities, and characteristics of the individual such as honesty, openmindedness, considerateness, etc.; and (b) type mental prerequisites to successful performance (behavior) of one sort or another, such as knowledges, skills, attitudes, ideals, interests, and appreciations. Each of these forms has its own peculiar advantages and disadvantages, educational and psychological. A composite approach has been employed herein as an examination of the statement of objectives will indicate.

[2] Such as, The Report of the Commission on the Social Studies of the American Historical Association.

[3] Cocke, Edith Herrin. *A Study of the Objective in Teaching American Government.* Master's Thesis, University of Wisconsin, 1936.

Rogers, Barton J. *The Objectives of Community Civics.* Master's Thesis, University of Wisconsin, 1936.

THE COLLECTION OF DATA

The data of the investigation here reported was collected by advanced graduate students pursuing the doctorate degree in education at the University of Wisconsin acting as special investigators. All of the teachers were visited, most of them a number of times. In the earlier investigations the tests to the pupils were administered by the teacher; in the later investigations the pupil tests were administered by trained investigators. The tests given to the teacher were given by the special investigators individually or in small groups. More detailed information relative to the methods of collection of data will be given in the special reports to follow.

THE CRITERION OF TEACHING EFFICIENCY EMPLOYED IN THE INVESTIGATION

The primary criterion of teaching efficiency employed in this investigation is that of the changes produced in pupils as measured by carefully validated tests of certain desired outcomes of instruction in the field of citizenship. The tests applied to the pupils have already been listed.[4] While no attempt will be made to explain here in detail how this criterion was developed, in general, the criterion consists of a composite of the change scores made upon the tests employed in this investigation. It will be recalled that unit tests were given at the beginning and end of each of two units of instruction and over-all tests given at the beginning and end of the school years. The over-all tests were designed to measure civic information, civic action, civic attitudes, civic beliefs, ability to organize research materials, and the application of generalizations to social studies events; the unit tests were designed to measure civic action, civic information, and civic attitudes. A supplementary record was taken of pupil activities arising from the work in the two units of instruction, but there seemed to be no very satisfactory method of employing these data, consisting as they did of qualitative statements, in the final criterion of teaching efficiency. The final criterion consisted therefore of a certain very restricted list of paper and pencil tests applied to the pupils as pre and final tests, and as is generally recognized today, such tests, constructed as carefully as they may be, are not without very definite limitations.

[4] Our concern in this investigation is with teacher effectiveness in the classroom. Important aspects of teaching efficiency, such as, teacher-teacher relationships, teacher-supervisory relationships, teacher-community relationships and extra-classroom teacher-pupil relationships are not under investigation.

The part of each change score attributable to the efforts of the teacher was considered to be the difference between the change scores secured from the application of the tests to the pupils and the score predicted for each pupil from a four-variable prediction equation based upon measures of the pupils' intelligence, socio-economic status, reading ability, and the pupils' pre-test score. The statistical procedures employed in making these calculations will be described in the special reports to follow. The other criteria employed in the investigation will also be described later. It is probably unnecessary to say that the investigators considered the criterion of teaching efficiency employed in the investigation a very important matter and gave it very careful consideration. The worth of all that follows depends in a very real measure upon the adequacy of the criterion of teaching efficiency employed in the investigation.

THE CONTROL OF FACTORS EFFECTING THE OUTCOMES OF INVESTIGATION

The investigators have attempted to control a number of factors thought to bear upon the outcomes of the investigation:

1. The subjects employed in the investigation were chosen from definite types of schools and homes, namely, 7th and 8th grade pupils enrolled in one room rural and state graded village schools, all engaged in the study of citizenship. Each sample has been carefully described and tested for homogeneity.

2. The teachers chosen for investigation were all chosen from non-departmentalized schools. The investigators could discover no easy way of discounting the effects of other teachers upon pupils in departmentalized schools.

3. An attempt was made to equalize the amount of time devoted to the study of citizenship. Each day's work was to consist of forty minutes in one class period or two periods of twenty minutes each. There was to be no home work. Where there were field trips and other activities these were to be employed in such a manner as not to increase the total time given to the work in this field.

4. Realizing that the equipment with which teachers would work might vary considerably, the investigators attempted to supply each teacher, upon request, with supplementary reading materials, equipment, and visual aids. A list of these materials was supplied each teacher.

5. Reference has already been made to the pre-tests and the tests of intelligence, socio-economic status, and reading ability employed in equating pupil ability.

THE STATISTICAL TREATMENT OF THE DATA

A variety of statistical procedures have been employed in the treatment of the data collected, such as, correlations between criterion scores and measures of teaching ability; item analyses of tests and other measures of teaching ability; analyses of variance; critical ratio studies of the differences between good and poor teachers; factor analyses and case studies. In general an attempt has been made to employ appropriate statistical procedure including both the better known standard procedures and those more recently developed in the statistical treatment of the data collected. The detailed procedures employed in this investigation will be described in the special reports to follow.

SOME VERY TENTATIVE CONCLUSIONS

While the analysis of all the data is not as yet completed the investigation has probably proceeded far enough to justify certain very tentative conclusions, first as to the methods of the investigation, and secondly, as to the nature and measurement of teaching efficiency. The following conclusions as to methods of research may be suggested:

(1) While extreme care has been exercised in building up the criteria of teaching efficiency employed in this investigation further work in this field should lead to the building up of better criteria. This will involve in the opinion of the writer, (a) a new study and re-statement of the purposes of citizenship instruction; and (b) new carefully validated, and more comprehensive measures of pupil growth and achievement in this area; as well as the refinement of other criteria.

(2) A more comprehensive measurement of teaching ability with more adequate measures in most areas.

(3) Better controls of the conditions under which the data are collected. While extreme care was exercised in the collection of the data employed in this investigation, the matter of controls needs further consideration. Refined statistical procedures are most essential in an investigation of this sort but they were never intended and cannot make up for the losses in information arising from poor data.

(4) This investigation will seem to some people very complex, but teaching is complex and in the writer's judgment investigations of teaching may have to be complex considering many things, each in relation to the other. The single variable type of research important as it is, even in the study of teaching efficiency, cannot take the place of semi-controlled multiple variable studies of the more complex problems

of teaching, particularly from a practitioner's point of view and the total process.

(5) This investigation is based in part upon a case study approach, but it seems to the writer that this approach must be extended. The potency or contribution of specific acts of the teacher must be considered in relation to situations that give rise to them; the consistency of behavior and its appropriateness would seem to relate too to the personal idiosyncrasies and personality of the individual teacher. Frequencies and statistical procedures ignoring this fact may not be meaningful and may possibly be misleading.

(6) There would appear to be need for long time genetic studies following individual teachers from the beginning of training or earlier through the training period and into service in which the relationships and growth of abilities essential to teaching efficiency are considered.

CONCLUSIONS DRAWN AS TO THE NATURE AND MEASUREMENT OF TEACHING EFFICIENCY

(1) It would seem apparent from the data examined that the teacher brings her whole self to teaching and not merely a knowledge of the subject to be taught or the methods of teaching. This conclusion seems to bear upon the perennial argument about whether good teachers are born or made. Certainly most of them are either born or made by the 18th or 20th year of school work.

(2) Personality is an important factor in teaching success. Teaching is or should be a very human undertaking. The emotional stability of the teacher; her social adjustments; her energy, vitality, and enthusiasm; her personal appearance and habits; the richness of her experience and background; skill in expression; and ability to work with others effectively all bear directly upon the effectiveness of the teacher. This is probably just another way of saying that the teacher brings her whole self to teaching.

(3) The measures of the various phases of teaching efficiency now available would appear for the most part to be quite inadequate, at least with reference to the criteria of teaching efficiency employed in this investigation. This conclusion seems to be supported by both the correlation studies which gave for the most part very low coefficients of correlation and the item analyses in which few measures seem to discriminate between good and poor teachers. This result may, however, be the product of many factors: errors of measurement; narrowness of the

range of talent measured; the small contribution made by individual test items to teaching as a whole; the failure to take into consideration the person-situation meaning of item, etc., and should be accepted with care. One suspects too from an examination of the instruments now commonly employed in the measurement of teaching efficiency that the items of which they are composed are quite artificial and very much removed from actual learning and teaching situations. Possibly test constructors may have to move closer to the scene of action.

(4) Of the measures employed in the investigation, intelligence as measured by the *American Council Psychological Examination,* knowledge of subject as measured by the *American Council Civics and Government tests* and interest in teaching as measured by *Yeager's Scale for measuring attitudes toward teachers and teaching profession* appear to possess somewhat higher validities than other tests employed in the investigation. It should be remembered, however, that these conclusions are very tentative. A more detailed report of the investigation will be given at a later date.

AN EFFORT TO EXTEND THE MEASUREMENT OF THE RESULTS OF SCHOOLING INTO THE SOCIAL AND ECONOMIC FIELDS*

HAROLD F. CLARK
Teachers College, Columbia University
and
MAURICE F. SEAY
University of Kentucky

Editor's note: One of the important outcomes of school instruction is the modification of behavior or action in some significant respect. The authors present a brief report of their efforts to extend measurement into the social and economic fields.

Do the schools really improve home membership or the use of leisure time? Or even health or vocational opportunity? It is easy enough to find statements based upon faith, saying that the schools improve living in these areas. It is practically impossible to find any experimental evidence that would be acceptable as far as scientific technique is concerned, establishing the fact that schools have improved life in these fields.

Many things in this world must be done on faith; perhaps the most important ones must. And for long into the future this situation will undoubtedly continue in regard to many of life's activities.

If you are asked whether a democracy is better than a dictatorship you may answer unhesitatingly, Yes. But that answer will be based far more upon faith and your set of values than any scientific evidence that is now available or is likely to be obtained soon.

Of course, there is no objection to providing education on faith. And there is no particular objection to having faith that education will produce certain important results. However, the question that is to be asked in this paper is, why doesn't education begin to measure some of these results rather than continue to ask that they be accepted on faith. A little more scientific evidence and a little less necessity for using faith probably would improve education, and undoubtedly would bring great additional support to education.

This paper will deal with the techniques involved in measuring one relatively simple but rather important social effect of education. Can the schools improve the diet of a community? If a hundred school people were asked this question probably ninety-nine of them would answer, Yes. And if they were asked to cite experimental evidence in this connection probably ninety-nine would say they knew of none.

* Paper presented at the St. Louis meeting of the American Educational Research Association, 1940.

It is practically impossible to find any clear cut experimental evidence showing what effect schooling has had upon the diet of a community. I am quite sure that the lack of this evidence is not due to the inherent difficulty of experimentation in the field. It seems to me that it is due more to the fact that most people still believe the purpose of education to be to improve reading, spelling, arithmetic, geography, or the knowledge of military and political history, or some other rather abstract formal subject. For a generation we have been theorizing that schools should improve health, community and home relationships, leisure and working life, and similar items. But we have been very slow in building a school system fitted to do these things or in setting up measuring devices to show whether we had done them.

At the present time, with the assistance of other agencies[1] and individuals, I am working on an experiment to learn whether it is possible for schools to change the diet of a community simply by teaching about better diet. In the particular communities chosen for this experiment two further issues are involved, which grow out of the fact that they are exceedingly low income communities. One problem is, will teaching about a good diet make it possible for the people to get a good diet? The second and perhaps more important issue is, how long will the instruction have to go on before the diet in the community is raised to an optimum level?

When work was begun on this experiment I had assumed that we could turn to the record of educational experimentation and find study after study along such lines. I had expected to find the whole process of experimental technique, check and control of communities all worked out. But imagine my surprise when I discovered there was almost no experimental evidence showing the effect of school instruction upon the diet of a community.

There are three reasons for this paper. First, to outline an experimental technique as a very minor contribution in this field. Second, to obtain suggestions from others for improving technique in the field of measuring the significant results of education. Third, and of greatest importance, to get the American Educational Research Association to devote time, effort and space to measuring the broad economic and social results of our schooling, instead of spending all their time and energy upon the relatively unimportant part of our educational system.

Choice of Community.—The actual plan of this experiment is as fol-

[1] The experiment is supported by a grant from the Sloan Foundation.

lows: An experimental community was chosen. The community is among the lowest income communities in the United States. A community of this type was chosen in order to learn whether teaching about diet in a school could raise even the poorest of our communities up to an optimum diet.

It was very important to find communities that would be fairly stable in population over relatively long periods of time. Perhaps one reason why there has been so little experimentation in the broader social and economic effects of education, is that such experimental changes come about very slowly. A program of teaching and experimenting may have to extend over many years in order to produce any appreciable results.

For instance, one community was chosen. A careful study showed that a large fraction of the population moved in and out of this community every few years. This, of course, would make it difficult or impossible to know exactly what was being measured in experiments carried on over a long period of time. For this reason the community was not left in the final experimental group.

Another factor of major importance is the coöperation of not only the local school authorities but the higher school authorities. Our experimental schools are largely one-room country schools. Naturally, it was necessary to have the full coöperation of the county superintendent and the blessings, at least, of the State Superintendent of Public Instruction. Otherwise a change in administration might upset the whole experiment. This, of course, is possible in any case.

Experimentation dealing with the broad social results of education, also requires the coöperation of the parents and of the entire community. It would be a waste of effort to start any experimentation in this field unless the community is thoroughly willing to coöperate.

Testing Program.—Our schools were tested very carefully to discover their status in each of the old line school subjects. A school system might improve the diet of the community by devoting most of its attention to the problem of diet. However, if arithmetic or history or geography suffered as a result many people would want to know how much they suffered. For this reason we tried to get a picture of the knowledge of the students in regard to each of the school subjects.

Perhaps more for purposes of general comparison than anything else the schools were also given a group of intelligence tests. Someone might argue that the experimental school improved the situation in the commu-

nity because it had unusually able children. At least the answer to that question should be available.

Children were also tested in regard to their knowledge of diet. In this case where the aim was to improve the diet of the community, it was very important for us to know the community practices in regard to diet. And for this reason a special study had to be made of not only the diet of the children but the food sources of the community, the dietary practices, etc. It would be quite impossible to measure the effect of the school upon improving the diet of the community unless one had an accurate picture of the diet at the beginning of the experimental work.

As an additional control, a careful health examination was made of all children. We had reason to think that there were some food deficiency diseases in some of the experimental schools. And in any case, it was an important secondary check to know the general health of the pupils. A careful physical record will be kept throughout the term of the experiment.

Probable Results on Old Line Subjects.—The most probable result of the experiment on the old line subjects is that just as much knowledge will be learned. Children will probably learn to read just as well or better if they work primarily on the problem of diet than if they work with abstract reading materials. We have no reason to think that the arithmetic problems involved in the problem of diet will not cause children to learn arithmetic more rapidly than if they study arithmetic in general. Even in the field of history the children will learn more rapidly if much of the material is organized around the history of food, how man has improved his implements to obtain food, etc., than if history is organized around kings and wars. It is quite true it will be a different kind of history, but there is no reason to think that the student will not know just as much history. Surely it is a better history!

Much the same position can be taken in regard to geography. There is every reason to think that the student's knowledge of the world, will be more realistic if a good part of that knowledge is directly connected with the geographical factors affecting the food supply in various places of the world, possibilities of making additions to diet by a great variety of means, and similar factors. For instance, it would be possible to discuss the great addition that the olive tree has made in the diet of the Mediterranean countries. That might lead to an investigation of nut trees in the mountainous parts of the United States as a supplement to diet. A long list of such factors will occur to anyone. Nothing is now

known that would lead one to expect the student to learn less geography under such a procedure. As a matter of fact we would expect the children to learn far more geography.

Control Schools.—One serious objection continually raised to any effort to measure the socio-economic results of schooling is that the effect of the schools cannot be distinguished from the effect of other social agencies. From casual investigation, this is true. In no case can the effects of schooling be measured absolutely and with perfect accuracy. To raise this objection, however, simply means that techniques must be found to deal with this difficulty. The scheme adopted in this experiment was to find control communities as nearly like the experimental communities as possible. The fact that our experimental schools were largely one-room country schools made it relatively simple in this particular instance. In a more complicated situation an effort would have to be made to see that the communities were as nearly alike as possible on all factors except the one to be tested.

Our next step was to pick control communities close to the experimental communities. And this, too, raised difficulties in that if the communities were too close there might be a direct carry-over from the experimental to the control community. Our control communities are usually thirty or forty miles from the experimental communities but not along the major lines of communication. The district where the schools are located is fairly rough, and the distance mentioned seemed great enough to prevent any large amount of direct transfer. The communities are close enough to be essentially alike so far as food habits, population, general customs, and standards of living are concerned.

Influence of Other Factors.—Someone will say that the effect of the radio may improve the diet, and we do not know how much change is due to knowledge from the radio and how much is due to knowledge from the school. Our answer would be that we have no reason to think the influence of the radio would not be the same in our control community and in our experimental community. This would hold true of the county farm agents, the home demonstration agent, the agricultural papers, or any one of a long list of other agencies which could be mentioned. Presumably, all of these influences are at work both in the control community and the experimental community. Of course, we may discover that all of these other influences are so powerful that the addition of the school working on the problem in the experimental community would not be noticeable. If this is true then it seems to us there is fair evi-

dence that the school is not the crucial factor affecting diet in the two situations concerned.

Content of the New Course of Study.—The new school program in the experimental communities is being built around the problem of improving the diet in the communities. It so happens that the diet in our experimental communities is quite inadequate from many standpoints. One of the very first lines of approach will be to raise the question in many places in the school program, what is an adequate diet? At all levels the school program will be attempting to throw light on this question. At frequent intervals discussion will be held on what can be done in this particular community to move from the present diet toward a more adequate diet. We hope that by this means we will get the pupils in those schools to make many specific suggestions as to actual changes to be made in the diet of the community.

We know, for instance, that no adequate effort is made to keep a supply of green vegetables through most of the winter months. By relatively slight changes in the processes of growing and storing, green vegetables could be supplied throughout practically the entire year. The evidence is already clear that there is a shortage of calcium in the diet of these communities. The students will be confronted with the problem of finding a cheap source of calcium for their diet. There seems to be an easy solution in sight. Some technicians have told us that the addition of soy beans to the diet of these communities would help the diet in many places. Will this be a feasible thing to accomplish through the schools? We certainly do not know. It should be enlightening to find out.

Specific Methods.—If the schools are to attempt to improve community life, obviously they must have materials dealing with the factor to be improved. This means that in our diet experiment we face the problem of providing an enormous amount of material at all grade levels on the problem of diet. This is a major problem. We have also become convinced that the older type of school concentrating largely upon verbal material is an inadequate solution to the problem. The new school attempting to improve economic conditions, will probably use far more visual aids of all kinds than has the school of the past. We need literally scores of new types of charts and pictures demonstrating new methods of diet, showing pictures of processes and pictures of results.

Generalizing from Results.—One would always hesitate to generalize from the results of any limited experimental process. Even if our results showed that the schools do cause greater improvement in diet prac-

tices in the experimental communities than in the control communities, we would still not be justified in saying that the schools could do the same any place in the world. If, however, in each of our several experimental communities greater improvement is shown than in each of the control communities, the evidence would be reasonably strong for communities of this general type. If in addition in various other sections of the United States, and under various other conditions similar experiments were carried on, and in most of the cases there were great gains due to the experimental procedure, one would have a strong case for saying that if schools work along the lines of the experiment improvement will result. Even a few hundred such experiments, strategically located, and scattered through a wide range of items, should do much to broaden the evidence in favor of education.

Perhaps one note of warning should be sounded at this point. There is every reason to think that such experimental techniques would require very long periods of time to carry out. It may take a generation or even longer to show the full results of such a program. In the past educational research has not been organized to carry on such long time experimental procedures. It is important that this defect be remedied.

There is one other warning that I should like to mention. So many of the experiments that are carried on in education obtain extensive equipment, highly trained teachers and a great staff of special supervisors. In setting up this little project we were most careful to keep exactly the same school houses, the same inadequate desks and equipment, and even the same broken window panes. We definitely rejected the suggestion that special teachers be put in the experimental schools. What we have tried to do is to set up a technique that would change the mechanics of the school as little as possible. Then if the evidence does show that a school program by working on diet can improve the diet of the community, it would be reasonable to expect other schools to change in the same direction if they wanted to make that type of change in the community.

The primary purpose of this attempt to sketch the procedure we have evolved is to obtain criticism of our own procedure and to encourage other persons to begin to experiment on the broad social and economic results of schooling. The only sound support of education in the years to come must rest on some such procedure.

USE OF INTEREST INVENTORIES FOR PERSONALITY STUDY

G. V. Sheviakov and Jean Friedberg
University of Chicago

Editor's note: Much attention has been given in recent years to personal and social adjustment problems. The authors describe their attempts to measure adjustment through the use of interest inventories.

In connection with the work of Evaluation in the Eight Year Study of the Progressive Education Association an attempt is being made to study personal and social adjustment through the use of interest inventories. The following presentation will be concerned with certain very general aspects of the study. Because the study is not yet completed and because any presentation of possible conclusions would be both premature and difficult within the scope of this article, we shall attempt to give here only a general description of the nature and purposes of this approach.[1]

Three interest inventories, each containing 200 statements of activities commonly carried on by young people, have been constructed to date: *Interest Index 8.2a,* which samples activities involved in the study of various school subjects, and *Interests and Activities 8.2b and 8.2c,* in which the items are chosen from extra-curricular activities and out-of-school situations. The responses to the activities in all of the inventories are in terms of *Like, Indifferent,* and *Dislike.* All three of the inventories must be given for a personality appraisal.

The activities in *Interest Index 8.2a* are classified into the following fourteen categories: Mathematics, Physical Sciences, Biology, Social Studies, Reading, English, Foreign Languages, Business Education, Music, Fine Arts, Home Arts, Industrial Arts, Sports, Manipulative. The activities in *Interests and Activities 8.2b and 8.2c* are classified into categories as follows:

Acceptance of Own Impulses. This category contains activities which are characteristically impulsive or involve sensory gratifications, though on the whole they are the socially sanctioned activities in these areas.

Severity with Oneself. This category contains activities which are characteristically self-regulatory and self-censoring.

[1] A fuller description of the development of the inventories and illustrative data may be obtained in the mimeographed bulletin, *Evaluation of Personal and Social Adjustment,* Progressive Education Association, Evaluation in the Eight Year Study, (Chicago: The University of Chicago, 1939), 65 pp. Paper presented at the St. Louis meeting of the American Educational Research Association, 1940.

Relationship with Family. This category contains activities carried on in the home, with other members of the family.

Identification with Others. Interest in other people, activities involving warm, intimate relationships with others.

Non-identification with Others. Activities presumably involving a desire to be outstanding, exclusive, a wish to be admired, etc.

Relationship with the Same Sex. Activities which are typically engaged in with members of the same sex only.

Relationship with the Opposite Sex. (1) Group activities involving the participation of both sexes, (2) Activities indicating interest in romance and courtship, (3) Activities indicating a vicarious interest in the opposite sex.

Solitary Activities.

Aggression. Activities representing more or less socially sanctioned hostile acts or attitudes toward other people.

Fantasy. Daydreaming, imagining, thinking about the future, etc.

Mystery. Activities involving curiosity about the occult and hidden, own and other people's personal secrets, etc.

Magic. Activities involving resorting to magical powers, belief in magic of thought, the use of amulets, etc.

Humor.

Preoccupation with Cleanliness. Activities involving avoidance of contamination, concern about purity, being immaculate, etc.

Methodical (Repetitive). Activities involving routinizing, ritualizing, collecting, saving, etc.

Life-Death-Universe. Activities involving thinking or wondering about the meaning of life and death, eternity, the origin and end of things, the stability and permanence of the universe, etc.

Out-of-School Activities. Social activities outside of school.

School Activities. Activities involving student organizations, clubs, school life, etc.

Dramatics. Activities involving theatre arts, the taking of various rôles, etc.

Leadership. Activities involving organizing others, impressing one's ideas upon others, etc.

Authority. Activities involving submission to and rebellion against authority.

USE OF INTERESTS FOR PERSONALITY STUDY

With reference to interests the point of view of the dynamic psychologists is taken—that *interests reflect certain basic aims and desires.* The emphasis is on the meaning of interests within the personality structure of the individual rather than on their relationship to outside standards or value judgments of the desirability of certain interests. For the purposes of this study specific interests in themselves are rather insignificant; although their cultivation through reconditioning may be expedient as a means of achieving a given educational aim, the emphasis is not on the presence or absence of interest within a certain field, but rather on its significance for the inference of the underlying urges and aims. Thus interest takes on the significance of an index of emotional tendencies and of the organization of the personality of the individual. Liking and disliking, acceptance and rejection of activities become significant as expressions of underlying drives and goals within the individual, and as indices of how the individual operates in handling them.

METHOD OF INTERPRETING THE INTEREST INVENTORIES

The scores give the percent of likes, indifferences, and dislikes to items in each category. In studying an individual student an arrangement is made of his rank order in the class in percent of like and dislike responses in each category. The necessity of examining the student's individual organization of interests as such and also with reference to the group will be more comprehensible if the concept of adjustment which underlies this study is presented first.

Desirable adjustment is thought of as a process of maturation and adaptation during which the individual is able to integrate successfully (i.e., without neurotic compromises or delinquent episodes) his primitive impulses and drives with those expectations and demands which are imposed upon him (with reference to his age, sex, social status, race, etc.) by the group to which he belongs.

Thus the interpretation is carried on from two points of view: the particular pattern of the student's responses, and the relationship of his responses to the dominant trends of the group as these are evidenced in the ranges and medians.

First, the common element is sought in the categories liked most of all by the student, those most disliked, and those to which he is most indifferent. This first broad analysis gives the picture of the large areas

where he is receptive and outgoing, of those where he is hostile or rejecting, and of those areas where he is neither—where he is passive or non-committal.

In this preliminary examination of responses the meaning of one particular category is dependent upon its position in the constellation of the individual's preferences. For instance, high likes in *Leadership* and in *Fantasy* and *Magic*, coupled with low likes in *School* and *Out-of-School Activities* may have a very different meaning from a configuration in which a large number of likes in *Leadership* is combined with a large number of likes in such areas as *School* and *Out-of-School Activities*, *Sports*, etc. The first may be expected to reflect a wishful interest in leadership, the second is more apt to reflect an actual functioning as a leader.

The fact that with some students the category scores are apt to reflect their wishes rather than represent their actual behaviors is an important one and should not be regarded as something which makes the data invalid. On the contrary, is not this gaining of insight into the inner mental life of the subject the most difficult but important part of the problem? Careful study of results thus far has indicated that if one watches for the inner consistency of the picture presented by a student, one learns to discover facts about his fantasy life and learns to single out his wishful responses.

A deeper analysis may be made by an examination of responses to individual items. In the analysis of specific responses the same approach is again used. First the items liked in each category, then the indifferences, and then the dislike are taken; and each time an attempt is made to single out the common elements which characterize or run through the given group of activities. It is this approach which helps to bring to light with considerable specificity and clarity each individual's system or pattern of responses.

Many kinds of individual patterns of responses which are not evident in the category scores become evident through these more detailed analyses. One student may, for instance, like all items in the category *Family*, except those with the word "parents" in them. Another may dislike all overt aggression activities such as those involving hitting, swearing, fighting, etc., but may like such indirect, more or less surreptitious activities as "annoying people who are scolding me by remaining perfectly cheerful."

SPECIAL FEATURES OF THE PRESENT APPROACH

The features of this appraisal technique which may prove valuable in the study of personality are: its indirectness, its flexibility, the fact that the interpretation aims to be descriptive, and the fact that it attempts to gain insight into the dynamics operating in the individual.

Most of the items in the inventories have been obtained from children's diary records of their daily activities. Other similar methods of obtaining statements were used. Whenever possible, youngsters' language was preserved in the questionnaires. It is felt that the inventories do not appear to the students to be obviously a personality test, and that therefore they do not arouse the anxieties which many such tests evoke. The items are less crucial, less obviously diagnostic, and as such, perhaps less threatening to the subject. It seems possible that the distortion in response which occurs when a limited number of evidently diagnostic questions are asked may be eliminated by using a larger number of less crucial items and considering the significance of the responses to these items as they relate to each other and as they relate to their common denominator—the category of which they are a part.

Perhaps the greatest advantage of the approach is its flexibility. The inventories do not attempt to discover whether the student does or does not fall into one of a group of patterns prearranged by the investigator. Rather they attempt to provide a field upon which, within certain limitations, the student may trace his own patterns. If one considers that in the present set of inventories every student is given an opportunity to express himself in three possible ways in reference to 600 statements of behaviors, the variety of all the possible configurations becomes rather impressive. This is of great importance since it affords an opportunity for each individual personality to paint its own profile on a screen with 1800 points of reference. The subjects are thought to reveal their various affective trends through the configuration and the interrelation of their responses.

The final picture is a description rather than a score. This picture is derived from the way in which the individual student reacts to a great many fields of activity: academic interests, sociable activities, and activities which indicate his attitude towards himself.

In linking these seemingly quite different fields, an attempt is made to discover the common elements which make the youngster's response to academic situations understandable in terms of the way in which his personality and goals are organized. In connection with this it is felt

that only too frequently there is a dichotomy in our concept of a personality. The thinking life of a student is thought of as a discrete, separate unit determined by his I.Q. and "special abilities" and unrelated to his tensions, ambitions, fears and goals. The approach outlined above aims to bring to light certain common trends in the individual which evidence themselves both through his academic interests and other activities. Should it become possible to give a classroom teacher an instrument which will enable her to relate the strivings and the goals of a student and the possible satisfaction of these goals to work on certain academic problems, the opportunity to make education truly meaningful and helpful to children will be increased greatly.

Although a number of cases analyzed so far have yielded very gratifying results, a positive statement regarding the outcome of the study is premature. Because of this the inventories are not yet available for general use. Any widespread use of these questionnaires must be postponed until more thorough validation has been achieved and more information about the exact meanings of the items to the people responding has been accumulated.

In the validation procedure several teachers are asked to rate and to describe the students on a Descriptive Trait Profile. After the interpretations of the responses have been made, they are compared with these descriptions by the teachers to see to what extent there is agreement or disagreement. A similar check with the descriptions of students prepared by specialists such as school psychologists and child psychiatrists is also being made. It is planned that in order to determine objectively the amount of agreement between the two descriptions, a group of impartial judges will be selected to give their ratings of the degree to which the interpretations coincide with the descriptions by people who know the students.

The inventories have already been given to some 1500 students ranging from the seventh grade through college. They will be given to 1500 more this spring. It is hoped that a statistical analysis of the responses both in terms of category scores and responses to individual items will yield some interesting material on sex and age differences and differences due to socio-economic grouping. Extensions of present studies of the reliability, consistency of categories, significance of differences in response, etc., will be conducted.

RESULTS FROM FACTOR ANALYSIS*

With Special Reference to "Primary Mental Abilities"

JOHN M. STALNAKER
Princeton University and College Entrance Examination Board

Editor's note: One of the new statistical tools that has come into general use within very recent years is that of factor analysis. The author sounds important warnings relative to its misuse.

"THE faithful workers in the vineyards," those actually working first-hand with test material, should, I think, be more active in considering the simple explanations, the reasons, and the facts concerning the popular movements in testing. We too should theorize and perhaps cast our eyes toward the future and engage in a little very dangerous prediction as to where these movements will eventually lead. This paper is an effort to consider in this way factor analysis as applied to aptitude tests. Today, what topic concerning testing is more talked about or more written about than is factor analysis? Educational statisticians, including testers with a propensity for mathematics, psychometricians, and lesser psychologists, and even a few well-trained genuine mathematical statisticians, are having their say in multitudinous articles and books. The complexities involved in understanding them are great for many of the men working in the field. The way is beset with difficulties—even with snares and traps. Some of the writers, one dares to suspect, glory in their abstruseness and strive with success to be not too comprehensible. Even if with effort one follows them, caution must be observed lest a by-path prove so attractive that the main road is forgotten.

The man working in test construction and test development but not especially competent or trained in certain aspects of higher mathematics—and recognizing this limitation—should perhaps not even dare to discuss a topic as complex and technical as factor analysis. May he not, nonetheless, legitimately raise certain questions and call attention to the basic tests and their nature—the tests which underlie all the grand conclusions regardless of the method of analysis used?

The mathematics of factor analysis others have expounded. One point about the mathematical aspect, however, may be worthy of repetition. Only by Hotelling's theory is it possible to determine the statistical significance of the factors. For other theories no standard errors are yet available so one cannot tell to just what degree the results are statistically

* A paper read on October 28, 1939, at a conference in New York City of the State Testing Leaders, under the sponsorship of the American Council on Education.

dependable—or just how many factors have statistical justification. It is possible that at this time error factors, constant or variable, are being assumed to be factors of significance. This absence of statistical checks is a serious limitation in the present theories. We must look to the future for the development of these standard errors.

Another serious mathematical limitation in the popular theories is that in most of them there is no unique factor pattern. One must rotate axes until factors are found which appear to make psychological sense. Obviously no one argues against results which make psychological sense. Unfortunately, however, what is psychological sense for one is not so sensible for another. The truth is that one, be he psychologist or otherwise, cannot look at a test item and tell exactly what fundamental ability it tests. Accordingly, objection can be raised to some extent to the rotating of axes until the data make psychological sense in the mind of one experimenter. An objective criterion for rotation would be helpful. Mr. Thurstone has made some steps in this direction with his theory of simple structure, but the theory is not basically mathematical and not beyond question. And when should we stop subdividing factors? What criteria can we apply so that we may know that at last we have the real pure factors and that no further subdivisions are possible?

If we ignore for the moment the mathematics of the theories, we may be able to focus our attention on certain psychological matters. Evidence widely collected suggests that a single score typical of the composite so-called intelligence test has its limitations. There is need of some sort of breakdown of the composite test. The application of factor analysis gives the desired several test scores. Now one must pause and reconsider. What possible inferences might be drawn from these test scores concerning the mind? It is desirable that our practices have a plausible theoretical justification. Is the mind made up of two or five or seven or ten basic and independent abilities? Probably not. More than likely the human being has an almost limitless number of abilities in the sense the term is now being used. These abilities are, of course, a result of both heredity and environment. It is a little foolhardy, or at least premature just now, to attempt to locate the currently postulated factors in the genes. The speed tests of the simple paper-and-pencil, objective, machine-scorable type may be limited and may ignore completely many abilities of importance, possibly even the chief abilities of significance. In any case, one must remember the factors are merely mathematical constructs—hypothetical creations which may not have any real counter-

part. One should also remember that actually factors are described in terms of tests, not tests in terms of factors. Even though factors are the result of mathematical rationalizations and exist only on paper, they may, nonetheless, point to results of practical value. It is to be hoped, however, that ardent factor analysts will not simplify the mind, even for the minds of the simple, to a greater degree than the phrenologists have done before them. In time the application of factor analysis to superb test data may yield some hints concerning functional entities, and from those hints theories about the behavior of the mind may naturally develop. But these developments lie far ahead in the dim and remote future. At the moment the psychologist has no accepted theory of the mind which is accounted for by a few factors. Such theories may be developed and gain acceptance.

To the educator tests have a magical meaning. Having quickly considered the mathematician and the psychologist, we may now deal with the chief user of test results. The educator wants more and more sound aptitude measures, as do all persons interested in tests. The more revealing the profile chart of aptitudes, the better. A genuinely valid subdivision of the present composite aptitude scores would enable him better to guide students and eventually to gain insight into the abilities called for in certain types of training. Valuable means of predicting success in professional training of various kinds could result. These outcomes are all contingent upon valid tests of significant factors—upon scores which will stand individual interpretation. Such factors have not yet been established; such tests have yet to be developed.

The test technician can rightfully step forward at this point and raise questions about the tests at present being used as a basis for theoretical conclusions. How can sound conclusions concerning functional unities—primary mental abilities or factors—be drawn from the analysis of poor test material? In the rush for factors of great importance, tests of dubious and limited value are used—short, highly speeded tests, in which the difficulty and validity of the individual items are usually ignored. If such tests are given to a heterogeneous group of students, the scores can be "factor analyzed" and certain factors "found," although the word "found" is not an accurate one. The fundamental condition here of relatively low correlations among short speeded tests dealing with what appear to be very different forms of content is not new. Indeed the more important problem, on the contrary, is to develop several tests, each using a different form, which yield very highly related scores.

It may be appropriate to digress for a consideration of one or two elementary but basic notions about test results—notions which should be reviewed from time to time. Generalizations based on the results of correlations of test scores are dependent on the adequacy of two rather distinct types of samples, and the importance of these samples cannot be overestimated. One universe from which a sample is drawn is the universe of human beings who take the tests. All human beings are not alike, fortunately, and different types of subjects may yield different results. Studies made at the laboratory of the College Entrance Examination Board in Princeton show that the selection of the population is of great importance in correlational studies dealing with test results. One correlation may be obtained if a specific heterogeneous group is used, a quite different result if the group is restricted to a given age level or divided according to sex, to location and size of the home town, to the type of school attended, to the economic level of the family, and to other factors. Some of these differences, but not all, may be accounted for by the usual random sampling fluctuations. The nature of the basic population must be known, and in so far as possible controlled, if results of fundamental significance concerning the functioning of the mind are to be obtained.

The other universe from which we sample is that of the test material. We assume a homogeneous universe, for example, of items dealing with the verbal factor, and one particular test samples a few such items. Here again we must exercise the usual precautions to see that our sample is adequate and representative.

With these considerations in mind, let us scrutinize the tests now being used. Thus far, only one set of tests based on factor analysis has appeared, but we may expect more. Mr. Thurstone, after many years of extensive theoretical and voluminous experimental work, has recently had published by the American Council on Education a set of tests for what he calls primary mental abilities. We are all indebted to him for making available to us a set of tests embodying the results of his extensive and stimulating research on the subject. In the Princeton laboratory of the College Entrance Examination Board a rather thorough study of a few aspects of these tests has been made. Mr. Tyler, the Chief Examiner at the University of Chicago, made available for the study the test papers of their freshman class.

The most striking characteristic of the tests, obvious from inspection, is that they are all short speed tests. Manual dexterity in operating a

pencil and speed of reaction are of unquestioned importance in almost all the tests. Speed is so important, and the tests are so brief, that slight errors in timing must assume major importance. The test results would differ greatly for various age groups. Adults, on the average, cannot work at the speed demanded—such as in the "same-opposites" test in which twenty items per minute are offered and then the response word may be the same or the opposite of the stimulus word. In using machine-scorable sheets some subjects, it has been found, blacken the space between the lines very carefully. Such a person will be found deficient in all primary mental abilities. There is evidence that a very different factor pattern would emerge if these tests were given without time limits.

Although speed is certainly of primary significance in these tests, no factor or primary mental ability of speed has been found. That it has not may be a reflection on or limitation of the method of analysis used.[1] Does it make psychological sense to ignore speed? The College Entrance Examination Board has been working for many years with a test of the verbal factor—one of the most significant factors and one of the easiest to measure thoroughly. In its test five subtests are used, the time allowed for the shortest being fifteen minutes; for the longest, thirty-five. Two hours of actual testing time are used. The test is definitely a timed test, but the correlation between a time form and a parallel form given under unlimited time is .94. The usual parallel-form reliability is around .90. The reason that time is not so important is that all items within a subtest are arranged in order of difficulty and the level of difficulty is so controlled that it fits the population for which the test has been developed. The results of this test are suitable for individual guidance and interpretation. Since the University of Chicago freshmen who took the tests of primary mental abilities also took this test, certain comparisons can be made. The correlation between the Thurstone primary mental ability score in the verbal factor (obtained by adding, according to his instructions, the scores on the two verbal tests) and the test of the College Entrance Examination Board was for this group .71. Mr. Thurstone estimates the correlation of his test with a test of the pure factor to be .68. If only one of the two tests which Mr. Thurstone uses for his verbal factor, the same-opposites test, is considered, its correlation with the longer test of the Board is .68. If the scores on only the first half

[1] A speed factor has been claimed by other investigators, *e.g.*, see Holzinger and Swineford, "The Stability of a Bi-Factor Solution," *Supplementary Educational Monograph*, University of Chicago, 1939.

of the same-opposites test are used, this correlation is raised slightly to .70. This latter method of scoring obviously gives a score less affected by the speed factor. In this case, cutting a test in half—using only half the items—seems to raise its validity—or at least not to lower it.

Let us scrutinize this same-opposites test further. It contains 500 words in its 100 items. Approximately 10 percent of the words appear twice in the five-minute test. Three words are used three times. One whole item is repeated with a change in but one distractor. The items, which are not arranged in order of difficulty, are chiefly easy items. The chances are one in four of getting a correct answer. The most difficult item was answered correctly by 46 percent of all the candidates who attempted it. Two thirds of the items were answered correctly by over 90 percent of all candidates attempting the item. Four of the items were correctly answered by every person attempting them. Of the remaining 96 items, six had a zero or *negative* biserial correlation with the total verbal composite score to which it contributed. If the longer test of the Board is used as a criterion, ten of the items are *negatively* correlated with it.

Since the verbal factor is well established and lends itself readily to exact and dependable measurement, it seemed worth looking into. The analysis suggests that the test can easily be materially improved. The question naturally arises, if it is radically improved, if the speed factor is removed, what then will be the result? Suppose all the other fifteen tests of this battery are likewise improved, what will the intercorrelation then be? And what factors will then be found?

The verbal factor is fairly well established. What about such a wobbly factor which has been given the name of deductive reasoning? The composite score representing it correlates .58 with another supposedly independent composite, induction. And yet two of the tests added to give the composite of deduction correlate .30 for the boys and .20 for the girls. Here there is less community among the tests added to give the factor composite than there is between factors! A detailed analysis was made of the items of the tests used to measure this factor, with similar results to those of the tests for the verbal factor.

Mr. S. S. Wilks, the consulting statistician of the College Entrance Examination Board, in looking over the study which was made of the results of the tests of primary mental abilities suggested a procedure equivalent to plotting the test scores in sixteen dimensional space for these students, letting each test be a dimension and each student be represented by a point. The principal axes of the resulting cluster of points

was then determined. This procedure determines what is commonly known as Hotelling's principal components. This was done for the scores of the boys. The result was that the points are stretched out mainly in one direction. The sixteen components are normalized so that their sum is 16. The first component is almost 5, the next is 2, then 1.5, 1, and so on down. Only the largest component was found to be significantly different statistically from the others. Mr. Wilks concludes from this evidence that the test scores of these 421 boys contain information relative to only one statistically significant factor.

To summarize: the mathematical theory underlying factor analysis techniques will doubtless be perfected in time so that factor analysis will be recognized as a powerful mathematical tool. Its application to the results of valid tests given to various but known types of populations may yield data of value for a better understanding of the functioning of the mind and for purposes of practical educational guidance. At the moment, however, the greatest need seems to be for vastly improved basic tests, tests developed with but one object in view—a thorough and dependable measure of certain types of ability. Such tests will require more time from the candidates; some of the tests may be more difficult to score; all of them will require great care in construction. Until such tests are used as a source for the primary data, positive conclusions from any system of factor analysis must be viewed as conjecture. To Mr. Thurstone we are all indebted for stimulating interest in this problem by offering the experimental edition of his tests, tests which he personally advocates be restricted at this time to experimental purposes and to group not to individual interpretation. One can but hope that a like amount of intelligence, energy, and industry will be devoted, by him and by others, to producing tests of the high quality necessary to make the results from mathematical analyses of test data of practical significance, not alone for basic theoretical studies, but also for the use of those who are merely "faithful workers in the vineyards."

SELECTIVE APPLICATION OF THURSTONE TEST

HAROLD S. TUTTLE
College of the City of New York

Editor's note: One of the important recent developments within the field of measurements is the construction of scales for the measurement of attitudes. The author offers important suggestions on their construction and use.

THE increasing range of application of the Thurstone technique for the construction of attitude tests necessitates a careful examination of the presuppositions underlying that technique. His demonstration that attitudes can be measured constitutes a significant contribution to the science of education. For the wide variety of topics upon which tests have been constructed the educational world is indeed indebted to the inventor of the technique. The ability to classify a group according to the degree of acceptance or rejection of a given belief is a genuine advantage to anyone who wishes to know or modify the beliefs of that group. The ability to express in mathematical units the change in degree of acceptance of a given belief on the part of the group after a period of training is a necessary factor in evaluating the method of training employed. Wherever there is value in discovering the relative strength of conviction on a given topic the tests built on the Thurstone technique render high service.

But the present wholesome trend toward emphasis on conduct and personality has opened up a myriad of new problems regarding the education of attitudes which are far more complex than the grouping of subjects or evaluating methods. The field of cultivating personality is fairly new. For every specific method of attitude training which has thus far been proposed and which might be studied experimentally there are a score of services needed for which methods have not yet been supplied.

At the present time policy making in the field of character education is a matter, not of testing a method already suggested, but rather of discovering many methods not yet invented for the building of desirable attitudes and motives. For policy making in the field of conduct training where new methods need to be devised to meet growingly conscious problems the Thurstone technique may in many cases be highly dangerous.

Implicit in the technique and explicit in its author's original statement is the assumption that an attitude is identical with an opinion. An opinion is the belief that a certain pattern will achieve a certain end.

But many educators believe that an attitude involves two distinct, or at least distinguishable, elements: (a) the desire for the attainment of the specified end; (b) the belief that the means in question will achieve that end.

One's attitude toward clothes is obviously not limited to the clothes themselves. It reaches on to some objective: one has a desire to keep cool in summer, one also has a desire to be approved by society. The clothes furnish the means for these desired ends. The weather is hot; one wishes to be as cool as possible. He believes that a linen suit will approximate that result as nearly as is feasible. But he must associate with social groups whose approval he desires. He fears that the linen will not long remain neat in appearance. His attitude toward linen is now unfavorable, though it had been favorable when he sought the goal of physical comfort. Obviously the man had conflicting attitudes toward the linen suit. Their algebraic sum may be expressed by a small negative score; but the significant facts are not revealed. For purposes of understanding the man's motives and determining a policy intended to lead him to action *it is necessary to know that he has a motive for comfort and a motive for social approval.*

There are, then, two distinguishable elements in all learning. One is the changing of one's *motives*—the changing of the goals which one desires; the other the equipping of one's intellect with data regarding the *means* by which the desired ends can be secured. Education of attitude involves thus the two corresponding processes—(a) the conditioning of one's interests so that his motives and desires are no longer what they were; and (b) the giving of information and guidance in organizing that information in systematic form. Desires may be produced by direct conditioning; logic always deals with means-to-ends, immediate or potential. An attitude, then, is the strength of one's tendency to utilize a particular pattern of behavior for the sake of attaining some desired end.

If we wish to know merely the degree of one's zeal for a certain *means*, then a measurement of his attitude toward that means is sufficient. For this the Thurstone test provides. But if we wish to control this desire to utilize a certain plan of action then it is important that we know the nature and relative strength of the two elements in the mental processes involved.

If the educator wishes to deal with the pupil who is not well adjusted in the school he must know whether the pupil feels little concern for the

welfare and reputation of the school or is ignorant of the effects of his conduct upon them. If the pupil lacks judgment he must be guided until he sees the effects of his conduct upon the group. If he lacks motive the educator must set up conditions that will cultivate a desire for the welfare of the school. If the pupil's motives are good but he lacks understanding of the effects of his conduct then punishment without conference and explanation will only tend to set up antagonism. If, however, he sees clearly the effects of his conduct but is not concerned for the welfare of the school, then reasoning is utterly without effect.

This principle holds true in every aspect of educating attitudes. Effective procedures are impossible unless the educator knows whether motives need to be strengthened or judgment needs to be guided.

This dual process in the education of attitudes the Thurstone technique wholly overlooks. "Opinions are the verbal expressions of attitudes." There is no distinction between judgment and motive.

This assumption of unity where there are really distinguishable factors may be illustrated in any field in which attitudes are involved. Take for example the controversy over the arms embargo and the cash and carry policy.

In a certain group of one hundred men let us assume that seventy-five want to keep the country out of war, while twenty-five wish the United States to engage in the war on the side of England and France. We thus start with two groups whose motives are diametrically opposed—who desire opposite and conflicting goals. Within the group which wishes to oppose war forty-five are convinced that the embargo on arms is the best means of accomplishing this result. We need not be concerned here with the logic by which they reached this conclusion nor with the soundness of their judgment. We are concerned here only with the fact that forty-five are convinced that the embargo will secure the result that they want, namely, to keep the United States at peace. Thirty of this same group, however, have become convinced that the cash and carry policy is least likely to lead us into the war. Again we need not be concerned with the logic by which they reached this conclusion. We merely know that thirty actually believe this and are acting upon their beliefs in attempting to determine public policy.

Of the twenty-five who wish the United States to participate in the war, fifteen have been convinced that the cash and carry policy is the surest way of involving us in the war. Again we need not be concerned with the logic or the soundness of this conclusion. And there is no lack

of realism in the picture; there are large numbers who do maintain this conviction. Of the twenty-five who wish the United States in the war, ten are convinced that the embargo is the surest way to draw the country into the war. This situation is also realistic. There are large numbers who so believe.

Now let us construct, according to the Thurstone technique, and apply to these hundred men, a test of attitude toward the cash and carry policy. What will we find? Forty-five of the one hundred will score positively on such a test. But who are the forty-five? Thirty of them are men who believe the cash and carry policy is the best means of preventing the United States from entering the war. Fifteen are men who believe that it is the surest way to lead the country into war. The motives of the two groups are diametrically opposed; and yet *the test will reveal no difference between them.* Fifty-five will score negatively in regard to the cash and carry policy. These fifty-five will be made up of forty-five who wish to keep peace and ten who favor participation in the war.

Now let us see what will happen when certain members of this group of one hundred are "educated." The accompanying chart will indicate the relation of attitude test scores to actual judgments and motives.

CHART SHOWING ACTUAL CHANGES IN BOTH GOAL AND JUDGMENT WITHOUT DETECTION BY "TEST OF ATTITUDES TOWARD CASH AND CARRY POLICY."

(Scores shown in boxes)

		First test	Changed judgment	Changed goal	Second test
		45 emb.	—5	—5	35 emb.
Antiwar	75	30 e.e.	+5	—5	30 e.e.
Pro-war	25	15 e.e.	—5	+5	15 e.e.
		10 emb.	+5	+5	20 emb.

By studying the arguments in the congressional record, ten are converted to the belief that the cash and carry policy will not tend to lead us into war. Of these five were anti-war and five were pro-war. When the test is repeated the score on the cash and carry policy will be diminished by ten. But has there been any change of motive? Not at all. The five who wish war will now oppose the cash and carry policy. The five who wish peace will now favor the cash and carry policy. There has been no change in motive. The only change is in the belief as to the effectiveness of a certain means for securing the desired goal.

Under emotional influences ten who had opposed war (divided five to five in their judgments) come to prefer war, without change of judgment as to the effects of the cash and carry policy.

Thus, twenty people have changed one or another aspect of their "attitudes" toward the cash and carry policy. Yet the second score gives no hint of such change, even as the first score revealed no evidence of conflicting motives.

Where educational policy making is involved, the Thurstone technique gives no indication as to the distribution of motive and judgment. For policy makers this instrument conceals the very elements which the educators must distinguish in order to determine the appropriate processes. Basic motives are concealed rather than brought to light. In situations in which the investigator is concerned to discover these motives the validity of attitude tests constructed on the Thurstone technique is zero.

EDITORIAL

THE GROWING COMPLEXITY OF MEASUREMENT

Professor E. L. Thorndike pointed out in one of his classical illustrations, a number of years ago,* that one can never measure an "elephant" but merely aspects, such as height, weight, pulling strength, color of hair, health, and so forth. Current efforts in measurement have come to reëmphasize this fact in their efforts to measure the child with reference to the purposes of education. Professor Thorndike also emphasized the difficulties involved in weighting test items or tests in the development of composite measures of human abilities, a problem the importance of which is becoming daily more apparent. Having spent some quarter of a century in attempts at the measurement of some of the simpler and more tangible aspects of education, measurement experts now turn their attention to the measurement of some of the more complex and less tangible aspects of education, and in such numbers that it is frequently necessary to place the results of a number of measures of child growth and achievement into some meaningful relationship.

Presented in this issue of the Journal is a series of papers each emphasizing the difficulties or importance of some particular aspect or application of measurement to the problems of education.

As one looks at the many present day activities in the field of measurements, there appear to be now in common use four fundamentally different approaches to the measurement of pupil growth and achievement, each with its own particular problems and assumptions. For lack of better descriptive terms, these approaches when considered from the point of view of the thing measured may be designated as (a) the type mental prerequisites approach wherein measurements are made of the knowledges, skills, attitudes, ideals, interests, and appreciations considered essential to efficient behavior; (b) the qualities of the personal approach, wherein measurements are made of qualities such as honesty,

* Thorndike, E. L. "The Nature, Purposes, and General Methods of Measurements of Educational Products," reported in the *Seventeenth Yearbook* of the National Society for the Study of Education, Part II, (Bloomington, Illinois: Public School Publishing Co.), p. 17.

industry, accuracy, sympathy, considerateness, etc., considered essential to success; (c) the overt behavior approach wherein inferences about the products of learning and experience are drawn from the study of observable pupil action; and (d) the measured or inferred effects approach, wherein evaluations are made from observed or measured effects such as, effects upon the social group of which one is a member, changes in social, political, and economic structure; changes in cultural standards, etc., resulting from individual action and participation. The delineation is not sharp but the trend appears generally recognizable in current efforts in the field of pupil measurements.

Each of these approaches has its own particular adherents and critics; its strong and weak points; its advantages and disadvantages. To illustrate this point, let us consider the overt behavior approach; I chose this because it is in some respects one of the newer approaches to evaluation. Suppose that one observed some forms of individual behavior and wishes to pass judgment upon it. How does one know in the first place that the behavior is not faked, and is not an honest representation of the individual hopes, beliefs, and ambitions? But passing this by, how does one know that the behavior observed was appropriate to the situation giving rise to it? There appear to be no universally good or bad acts as some would seem to suppose from current attempts to state the objective of education in terms of pupil behavior, except as behavior is defined so broadly as to include everything else. The acts of persons are probably of varying universality and appropriateness. When one speaks of acts, rather than principles of behavior, there are very few acts ordinarily considered good or appropriate that are not at some time bad or inappropriate. The appropriateness of an act is judged ultimately by its effects and this would seem to emphasize the joint importance of the inferred or measured effects approach to measurement. The behavior approach is chosen for criticism merely because its shortcomings are less well known. Each of these approaches has its own particular advantages and disadvantages and possibly more attention should be given to the study of these through the examination of the logic of each and the study of the consistency of the results secured from measurements made from these several points of view, with the hope that new insights may be had into the organization of human abilities and more effective evaluation of education outcomes obtained.

A. S. Barr

REVIEWS

Address all communications relative to book reviews to Editor Kai Jensen, University of Wisconsin, Madison, Wisconsin.

TIEGS, ERNEST W. *Tests and Measurements in the Improvement of Learning.* (Boston: Houghton Mifflin Company, 1939), pp. xxi + 490. $2.75.

This volume attempts to cover the field of educational measurement in a manner that will be practically useful to teachers. In general the author's viewpoint is sound, his presentation direct and clear, and his selection of material appropriate for the purpose he had in mind. However, the breadth of his purpose is such as to render it impossible to cover adequately measurement at all school levels in a single volume of ordinary length. The reviewer believes that it is preferable to treat either one school level or one large phase of the movement more fully. In his effort to do the other, Tiegs has left a number of discussions, among which are those dealing with certain types of tests, the use of results from physical examinations, final conclusions as to a no-failing program and as to marking, decidedly incomplete.

Among the commendable features of the book are the large number of figures and tables which help in clearness and forcefulness of presentation, the excellent statement of the need for better measurements, the brief history of the movement which might, however, well be longer, the discussion of the fallacy of the passing mark, and the concluding chapter with its questions frequently asked by students and the answers thereto. Some of the latter, however, are weak. The answer to Question 8 on p. 466 overlooks the fact that in many cases we do wish to measure rate of work. That to Question 9 on p. 470 defines scaling as scientific weighting whereas scaling is not necessarily accompanied by weighting at all. In dealing with Questions 5 and 6 on p. 475, Tiegs much underrates the value found in the use of both intelligence tests and pupils' ratings for the purpose mentioned. Unfortunately, the index both lacks many entries it should contain and is not arranged for most convenient use, hence renders it difficult to secure full value of the contents. Many of the examples of elements and portions of tests are not in good form although the implication is that they may be taken as models. They include errors in English, unsatisfactory directions, and features that make them more difficult to take and to score than is necessary.

The volume contains a number of statements that are either definitely in error or imply erroneous conclusions. On p. 21 it is implied that a norm must be an average, which is not true, as Tiegs himself points out elsewhere. The P.E. of an I.Q. from an individual test is given as 5 points whereas for the revised Stanford it is much less than that. Contrary to the statement on p. 82, there has been evidence reported that the arrangement of possible answers to true-false tests affects pupils' responses. On p. 241 it is stated that Larm and Odell have devised over one hundred English tests

whereas the correct number is about twenty-five. There are both misleading and definitely incorrect statements in the discussion of the accomplishment ratio. The index of reliability does not measure validity directly, as is implied on p. 386, but only indirectly through measuring reliability. Instructional costs per student are said to range from $75 to $300 per year, whereas the median for the country as a whole is below $75. The statement concerning test scoring machines seems to imply lack of acquaintance with their development in the last few years.

There are a few statistical errors. In two or three places the ranges of distributions are incorrectly given. In general the discussions of the significance of differences are based upon a conception of a sharply dividing critical point below which they are not significant and above which they are. This is a very misleading interpretation. The values computed for r_t and r_w on p. 382 are incorrect. In Table 38, x and y are reversed in comparison with conventional usage and fx is used with two different meanings. Several unusual rather than the commonly employed symbols, particularly subscripts, are used.

There are a number of minor omissions. Attitude and effort should be mentioned on pp. 17 and 18. The subdivisions of II on p. 23 apply to I as well and the general outline should include essay tests. Also the fact that T-scores are based upon twelve-year-old pupils is not stated. No mention is made of the use of anecdotal records.

Any volume contains some typographical and similar errors and this is at least as good as most in this respect except for an unusual number of misspelled proper names. "Ayres" is misspelled at least fifteen times. Without making any effort to proof-read the author noticed a dozen others, some more than once. There also are a few language errors of which the most noticeable are the apparently indiscriminate use of "youth" and "youths" as plural, "data" occasionally as singular, and "above" as an adjective.

Points of view are of course largely subjective but several are expressed with which the reviewer feels he must disagree. Chief among these are that the value of the I.Q. for individual use seems seriously underestimated, as also is that of teacher-rating and a single index of teaching efficiency, whereas the low correlation of intelligence and scholastic success, on the other hand, appears over-emphasized, and likewise the number of experimental data reported in connection with ability grouping, as compared with that bearing on other questions.

C. W. ODELL
University of Illinois

ADE, LESTER K. *Meeting the Needs of the Mentally Retarded.* Bulletin No. 420 (Harrisburg: Pennsylvania Department of Public Instruction, 1939), pp. ix + 168.

This bulletin, prepared by Dr. T. Ernest Newland and Miss Edna M. Kugler, replaces the earlier publication, *Organization and Administration of Special Education Classes for the Orthogenic Backward*, put out in 1935. As in the earlier work, major attention is given to the problems of teachers in small cities and semi-rural areas; and such matters as the philosophy of Special Education, the standards for special classes, the course of study, the methods, and the qualifications for special class teachers are treated at length.

Under *Equipment and Supplies* there is very valuable material on handicrafts in special classes; and the fine summary of reading techniques is retained. Much attention is paid to the curricular core

plan; and three illustrative units are outlined in detail. One of these on *Vocations and Occupations* could be very useful to teachers having children about ready to leave school.

RICHARD H. HUNGERFORD
Department of Special Education
Detroit Public Schools

BROOM, M. E. *Educational Measurements in the Elementary School* (New York: McGraw-Hill Company, 1939), pp. x + 345. $3.00.

Broom's new book on Educational Measurement in the Elementary school is definitely prepared for the use of beginning classes. Its organization lends itself to class work very easily. The introductory chapter furnishes the student with an idea of the historical development of measurement and many of its uses in the schools of today. The next three chapters supply a background of terminology and statistics. Then follow two chapters on the problems of the teacher-made objective examinations. A knowledge of the standardized tests is given in chapters dealing with tests in language arts, arithmetic, social studies, and a chapters on other subjects. Two chapters deal with the nature of intelligence and mental measurements, and his final chapter is on uses of educational measurement.

The volume can be commended from a teaching standpoint for its concise presentation, the short but carefully chosen references, and the many helpful exercises at the end of the chapters. A good course in measurement requires the student to develop a large number of skills. Most books have paid too little attention to the type of instructional exercises needed and have hoped that each teacher will supply his own. Broom goes a long way to correct this error. Where a class uses the text for background and the questions and exercises for study and discussion in class, the student would have an excellent idea of the measurement problems covered by the book. The text and the exercises supplement each other and it will not be sufficient to use one of them alone.

The standardized tests which have been selected for description are among the better recent tests. The descriptions are sufficient to give the student an idea of the type of instrument and some of its uses. It is always a temptation to include too long examples, but this has been avoided.

The problem of selecting material to be included in a course in measurement is a very real one for the field has grown to be so large. Each person has his own preference for materials to be included and the ways in which such materials should be organized. The reviewer would wish to include in a course on the elementary school level material dealing with attempts to study more carefully the personality of the child, the techniques which are developing of accumulating information on a pupil over a long period of time, and some of the newer attempts at evaluation.

In some cases where obviously it was not practical to include a complete treatment of a topic it might have been better to omit it entirely. Examples of such points are the treatment of the constancy of the I.Q. and the scoring of the rearrangement test.

A table of squares and square roots in the appendix will prove helpful to the student. The beginning student will find the volume helpful, and teachers of measurement courses will find the volume usable.

J. MURRAY LEE
University of Wisconsin

TRELOAR, ALAN E. *Elements of Statistical Reasoning* (New York: John Wiley & Sons, Inc., 1939), pp. xi + 261. $3.25.

The aim of the book is well stated in the preface. "Those who seek a compendium of statistical techniques, an array of formulas from which one or another may be culled to act as a mill for grinding fine flour from crude grain, by merely turning a handle, should not look further in these pages. What is written herein is intended for those who wish to reason carefully not merely to imitate." The book deals almost wholly with large sample theory. However, the author states that he may complete his original plan, by giving a critical analysis of small-sample technique in another volume.

Elementary algebra and some knowledge of analytic geometry is assumed, but the calculus is not used in any of the derivations. For example, the slope and intercept of the straight line of best fit are derived by a simple algebraic procedure which determines the minimizing value of the parameters by segregating them in a squared term and setting it equal to zero. The conventional topics are covered—central tendency, variability, normal curve, Chi square, and sampling errors of means, correlation coefficients and proportions. The careful and accurate presentation of sampling errors of means, differences between means, and correlation coefficients, as well as the clear statement of the logic underlying Chi square are especially noteworthy although some persons would probably disagree with the statement that "assigning significance where none really exists . . . is likely to be more serious . . . (than erroneously claiming insignificance)." The student is given a clear understanding of the difference between the "sample" on the one hand and the "supply" or "universe sampled" on the other, and of the relation between the parameters of the supply and the statistics of the sample which we must use in most cases as the only available estimate of the supply parameters. In short, while the mathematics used is elementary, the level of approach and the explanation of the logic involved are advanced.

There is interesting historical orientation throughout the book emphasizing the relative recency of much statistical development, e.g., the probable error of a proportion was not available to Mendel to aid him in interpreting his experiments on heredity.

The various statistical techniques explained are illustrated with a variety of data, such as a frequency distribution of 500 measurements of the width of a spectral band of light, determinations by two chemists of the protein content of the same flour sample, the dockage assessment on wheat by the state grain inspector compared with the assessment by the mill laboratory, the relationship between calcium intake and pyorrhea, and the effect of vaccination on recovery from smallpox. Aside from the problems worked out in the text, there are no arrays of special exercises in the book. Throughout, the text is liberally illustrated with graphs of the various equations introduced—showing the effect of ρ and N on the sampling distribution of r, the change in the distribution of Chi square with n, etc. The author thus follows the injunction from Balton given on the dedication page, "It is always well to retain a clear geometric view of the facts when we are dealing with statistical problems which abound with dangerous pitfalls, easily overlooked by the unwary, while they are cantering gaily along upon their arithmetic."

The book was developed by the author in teaching classes composed principally of graduate students. At this level of scientific maturity, the text can probably be used as an introductory statistics text. For less mature students in some fields it may be necessary to use it for a second

course in statistics, or to use it in conjunction with a more elementary presentation.

HAROLD GULLIKSEN
University of Chicago

WEBB, L. W., AND SHOTWELL, ANNA MARKT. *Testing in the Elementary School* (New York: Farrar and Rinehart, Inc., 1939), pp. xix + 407. $2.75.

In the field of educational measurements, as in most other, authors are faced with a choice between attempting to cover the whole area embraced by the term and dealing with a more or less narrow portion thereof, and also between being rather highly critical and the reverse. In this case, the authors have wisely narrowed the field and have assumed a point of view that is not highly critical. The book appears to be best suited for elementary classroom teachers of little or no previous training along the line covered who are primarily interested in practical classroom testing and the use of the results obtained. For such individuals this volume should offer much that is helpful and not overly difficult to master. If they wish to enlarge upon the content of the book, the many references will greatly assist them, but in some cases those given are not so up-to-date as they might well be.

The volume is almost entirely devoted to the use of standard tests in the school subjects and general intelligence. Most teachers make a greater use of informal tests which they construct than of standard tests. Personality measurement is receiving increasing emphasis. In this volume, however, informal tests are accorded little attention and personality measurements none at all. School marks constitute another topic that is decidedly neglected. A number of terms common enough in educational literature that teachers who read anything at all in that line are almost certain to encounter are omitted. Moreover, opinions and statements are too often quoted with little or no comment as to whether they should be accepted or not.

The general style is clear and understandable and the number of errors in English, spelling, and so forth small in comparison with some similar books. The effort to simplify material has occasionally resulted in statements that are not strictly accurate, perhaps through the omission of exceptions and qualifications. In other instances errors appear without appearing to have this origin. As examples of misleading or incorrect statements, it is implied that a norm must be an average; percentile ranks of 0 or 100 are mentioned, although such ranks are not statistically possible; the desirability of tests containing material of 50 percent difficulty is implied to be universal; it is stated that there is only one test of musical appreciation; also that there are "far fewer tests of art than of any other school subjects;" Monroe is given credit for developing the *B*-score; and the distinction between test and scale appears to be based on the definition of a scale as a scaled test, which is a much too limited one. The selection of group intelligence tests discussed in the text is not so thoroughly in accord with general use and opinion as it might well be and in the instances of a few of the school subjects tests worth mentioning have been omitted.

On the whole, this volume appears to deserve rank among the most helpful treatises on practical elementary school measurement, but may lead those for whom it is best suited to a degree of over-confidence in measurement techniques and an impression of over-simplification of the field covered.

C. W. ODELL
University of Illinois

RESEARCH ABSTRACTS and BIBLIOGRAPHIES

Address all communications relative to research abstracts and bibliographies to A. S. Barr, University of Wisconsin, Madison, Wisconsin.

BULLETINS

CARR, WILLIAM G., AND OTHERS. *The Status of the Teaching Profession*, National Education Association Research Bulletin, XVIII, No. 2 (Washington: Research Division, National Education Association, March, 1940), 79 pp.

This bulletin includes information on the following phases of the status of classroom teachers: general composition of the teaching population; educational preparation and certification; professional experience and mobility; salaries and economic condition; professional load; provisions for tenure; provisions for retirement; professional associations of teachers; mental and physical health; and academic freedom.

GOOD, CARTER V., JOHNSON, PALMER O., SEARS, JESSE B., AND OTHERS. "Methods of Research in Education," *Review of Educational Research*, IX, No. 5 (Washington: American Educational Research Association, December, 1939), pp. 449-646.

This issue of the *Review* is a summary of the studies reported in preceding issues and includes five types of material: new developments in research technics; evaluative studies of research technics; new types of applications of given technics; illustrations of current uses of technics; and notes on needed research.

GRAHAM, BEN G., ET AL. "Catalog of Aids to Perceptual Learning," *Pittsburgh Schools*, XIV, No. 2 (Pittsburgh: Board of Public Education, November-December, 1939), pp. 46-93.

This bulletin contains a list of materials consisting of silent and sound motion pictures, lantern slides, exhibits, etc. which are available to classroom teachers.

HARGRAVE, HAROLD, AND OTHERS. *Proceedings of the Fifth Annual Guidance Conference Held at Purdue University November 17 and 18, 1939*, Studies in Higher Education XXXVII (Lafayette, Indiana: Division of Educational Reference, Purdue University, March, 1940), 106 pp.

This bulletin contains the addresses presented at the Fifth Annual Purdue University Guidance Conference including: the LaPort guidance plan; the guidance program at Culver Military Academy; the guidance program in the Terre Haute city schools; the holding power of the Indianapolis high schools; guidance in relation to human needs; the evaluation of guidance practices; the Purdue Psychological Clinic; the IQ in a nutshell; problems of rural guidance; a rural guidance program; and regional guidance conferences in Indiana.

HOLMSTEDT, RALEIGH WARREN. *State Control of Public School Finance*, Bulletin of the School of Education, Indiana University, XVI, No. 2 (Bloomington, Indiana: Bureau of Coöpera-

tive Research, March, 1940), 71 pp.

The author presents the results of a study of state control of public school finance and discusses criteria for evaluating practices and procedures in such control.

LEWIN, KURT, LIPPITT, RONALD, AND ESCALONA, SIBYLLE KORSCH. *Studies in Topological and Vector Psychology I*, Studies in Child Welfare, XVI, No. 3 (Iowa City, Iowa: University of Iowa Press, 1940), 307 pp.

The authors discuss formalization and progress in psychology, an experimental study of the effect of democratic and authoritarian group atmospheres, and the effect of success and failure upon the level of aspiration and behavior in maniac-depressive psychoses.

MEAD, A. R., AND OTHERS. *Coöperative Study of Six Laboratory Schools*, etc., Bulletin Nos. 2, 12-16 (Gainesville, Florida: Bureau of Educational Research, University of Florida, January, 1940), Total pages, 115. Mimeographed.

These bulletins contain reports on studies conducted under the direction of the University of Florida Bureau of Educational Research, and include: coöperative study and work of six laboratory schools, 1938-39; academic grades of P. K. Yonge Laboratory School graduates in colleges and universities; Evangeline, an inclusive school project in the P. K. Yonge Laboratory School, 1934-35; vocational agriculture in the P. K. Yonge Laboratory School, and its values; small classes in Florida high schools; and continuous service of teachers and principals in Florida.

MILLER, L. PAUL. *State Regulation of Entrance into Occupations in the State of New York*, Abstract of Thesis (New York: School of Education, New York University, 1939), pp. 61-66.

The author presents the results of a study of state legislation in the state of New York which has placed requirements of personal qualifications upon individuals for legal entrance into certain occupations in the state.

REEVES, FLOYD W., CHMN. The Advisory Committee on Education, Staff Studies Nos. 1, 3-7, 9-12, 14-19, Washington: Superintendent of Documents, U. S. Government Printing Office, 1939.

The Advisory Committee on Education series includes the following bulletins: No. 1, Smith, Payson, Wright, Frank W., et al. *Education in the Forty-Eight States*, 199 pp.; No. 3, Frederic, Katherine A. *State Personnel Administration: With Special Reference to Departments of Education*, 271 pp.; No. 4, Heer, Clarence. *Federal Aid and the Tax Problem*, 101 pp.; No. 5, Mort, Paul R., Lawler, Eugene S., et al. *Principles and Methods of Distributing Federal Aid for Education*, 99 pp.; No. 6, Edwards, Newton, and Richey, Herman G. *The Extent of Equalization Secured Through State School Funds*, 55 pp.; No. 7, Hamilton, Robert R. *Selected Legal Problems in Providing Federal Aid for Education*, 71 pp.; No. 9, Blauch, Lloyd E. *Vocational Rehabilitation of the Physically Disabled*, 101 pp.; No. 10, Works, George A., and Morgan, Barton. *The Land-Grant Colleges*, 141 pp.; No. 11, Joeckel, Carleton B. *Library Service*, 107 pp.; No. 12, Wilkerson, Doxey A. *Special Problems of Negro Education*, 171 pp.; No. 14, Campbell, Doak S., Bair, Frederick H., and Harvey, Oswald L. *Educational Activities of the Works Progress Administration*, 185 pp.; No. 15, Blauch, Lloyd E., and Powers, J. Orin. *Public Education in the District of Columbia*, 99 pp.; No. 16, Blauch, Lloyd E. *Public Education in the*

Territories and Outlying Possessions, 243 pp.; No. 17, Blauch, Lloyd E., and Iverson, William L. *Education of Children on Federal Reservations*, 145 pp.; No. 18, Blauch, Lloyd E. *Educational Service for Indians*, 137 pp.; No. 19, Judd, Charles H. *Research in the United States Office of Education*, 133 pp.

RUCH, GILES M., AND SEGEL, DAVID. *Minimum Essentials of the Individual Inventory in Guidance*, Vocational Division Bulletin No. 202 (Washington: Office of Education, United States Department of Interior, 1940), 83 pp.

The authors discuss the importance of the individual inventory in guidance, essentials of the individual inventory, the value for guidance of items in the individual inventory, the determination of aptitudes, selection of tests, and selected tests with special reference to guidance.

SARBIN, T. R., EDITOR. *Proceedings of the Institute on Placement Services in Colleges and Universities* (Minneapolis: Center for Continuation Study, University of Minnesota, 1940), 52 pp. Mimeographed.

This bulletin contains a summary of the discussions at the Institute held on November 2, 3, 4, 1939, including recruitment in private employment, recruitment in public employment, training for employment in school, induction of college graduates into business and industry, qualifications and duties of the recruitment officer, and college placement procedures.

SEAY, MAURICE F., DIRECTOR. *Proceedings of the Sixteenth Annual Educational Conference and the Fifth Annual Meeting of the Kentucky Association of Colleges and Secondary Schools*, Bulletin of the Bureau of School Service, XII, No. 2 (Lexington, Kentucky: College of Education, University of Kentucky, December, 1939), 242 pp.

This bulletin contains the manuscripts of addresses and reports of the Sixteenth Annual Educational Conference and the Fifth Annual Meeting of the Kentucky Association of Colleges and Secondary Schools held at the University of Kentucky on October 27 and 28, 1939.

WEISS, MARTIN G., AND COX, GERTRUDE M. *Balanced Incomplete Block and Lattice Square Designs for Testing Yield Differences Among Large Numbers of Soybean Varieties*, Research Bulletin 257 (Ames, Iowa: Iowa State College of Agriculture and Mechanic Arts, April, 1939), pp. 293-316.

The authors present the statistical features of the designs of two experiments which are adapted to testing large numbers of varieties on variable soil.

WRIGHTSTONE, J. WAYNE, AND OTHERS. *Educational Research Bulletin*, XIX, No. 6 (Columbus, Ohio: Bureau of Educational Research, Ohio State University, March 13, 1940), pp. 149-186.

The authors present articles reviewing the Report of the New York Regents' Inquiry and the coöperative Study of Secondary School Standards.

EDUCATIONAL ADMINISTRATION AND SUPERVISION

TRAXLER, ARTHUR E. "Planning and Administering a Testing Program," *The School Review*, XLVIII, No. 4 (April, 1940), pp. 253-267.

The author outlines and discusses a program for testing school pupils.

PSYCHOLOGY OF LEARNING AND TEACHING

GINSBERG, WALTER. "How Helpful Are Shakespeare Recordings?" *The English Journal*, XXIX, No. 4 (April, 1940), pp. 289-300.

The author presents the results of an investigation of the effectiveness of recordings in the teaching of literature from which he concludes that the recordings proved a valuable aid to the appreciation of Shakespeare in the classroom.

PUPIL GUIDANCE, WELFARE AND DEVELOPMENT

DOLL, EDGAR A. "Some Things We Know in Clinical Psychology," *The Journal of Applied Psychology*, XXIV, No. 1 (February, 1940), pp. 20-26.

The author discusses some things generally agreed upon in the field of clinical psychology.

STATISTICS, MEASUREMENTS AND SCIENTIFIC TECHNIQUES

PRESCOTT, A. C., AND GARRETSON, O. K. "Teachers' Estimates and Success in College," *The School Review*, XLVIII, No. 4 (April, 1940), pp. 278-284.

The authors present data from which they conclude that rating by high school teachers of students' probable college success is more accurate than prediction by measures now commonly used for the purpose.

ROSTKER, LEON E. "The Measurement and Prediction of Teaching Ability," Reprinted from *School and Society*, LI, No. 1306 (January 6, 1940), pp. 30-32.

The author presents the results of an experimental study of the traits associated with teaching ability from which he concludes that a battery of objective measuring devices is useful for selecting and promoting teachers, but that rating scales are of little value.

NEW TESTS

BARR, A. S. *Teacher's Performance Record* (Bloomington, Illinois: Public School Publishing Company, 1940), 8 pp.

The author provides a means of evaluating the teachers performance in the classroom with reference to principles and conditions with separate scores on judgment in the choice of techniques, skill in the use of techniques, personal fitness, inferred results and general merit.

ROSANDER, A. C. *Social Attitude Scales* (New York: Association Press), each form, 1 p.

These scales, including the *Social Justice Scale*, the *Federal Constitution Scale*, and the *Democracy Scale*, each have two forms constructed on the Thurstone technique with simplified scoring.

WASHBURNE, JOHN N. *Washburne Social-Adjustment Inventory* (Yonkers-on-Hudson, N. Y.: World Book Company, 1940), 8 pp.

This test is for use in determining the degree of social and emotional adjustment of individuals over 12 years of age and includes subtests on truthfulness, happiness, alienation, sympathy, purpose, impulse-judgment, control, and wishes.

WORKBOOKS

STOGDILL, EMILY L., AND HERNDON, AUDELL. *Objective Personality Study*, A Workbook in Applied Mental Hygiene (New York: Longmans, Green and Company), 106 pp.

This workbook is designed for use by the college student in applying mental hygiene principles to the adjustment of the student himself.

RESEARCH NEWS and COMMUNICATIONS

Address all research news and communications to Carter V. Good, Teachers College, University of Cincinnati, Cincinnati, Ohio.

Genetic Studies of Genius.—Under the direction of Lewis M. Terman, the Stanford studies of gifted children are being carried forward through follow-up investigations. This group was first followed up in 1928, at which time (six years after the subjects were first located) the majority were given additional tests of intelligence and of one or more personality traits. Two field assistants made contacts with most of the subjects and their parents. The results were published in Volume III of *Genetic Studies of Genius.*

A second follow-up was made in 1936-37, this time chiefly through information blanks filled out by the subjects and their parents. The results of that follow-up have been summarized briefly and published in the 1940 Yearbook of the National Society for the Study of Education.

With the aid of five research associates Terman is now engaged in a thoroughgoing follow-up financed by a grant of $20,000 from the Carnegie Corporation. Field workers have been holding conferences with subjects and their parents, and have given the subjects and their spouses one intelligence test, two personality tests, and an interest test. Offspring of the subjects to the number of 300 or more have been tested. Extensive information is being secured on the marital adjustments of those subjects who are married, which is more than half of the total 1400.

The chief value of these genetic studies over a long period of time is in the longitudinal view of the development of the subjects. No adequate information had ever before been secured on how gifted children turn out. Ultimately the data will not only answer that question, but will throw light on why the individual subjects turn out as they do. Such longitudinal studies are, of course, valuable not only for gifted groups, but also for normal or defective subjects.

Secondary Curriculum.—The Commission on the Secondary School Curriculum of the Progressive Education Association, under the direction of V. T. Thayer, is in the process of publishing the volumes which represent the results of its work. *Reorganizing Secondary Education* appeared in the fall of 1939. It presents the general philosophy and point of view of the Commission and of its various committees. *Science in General Education* appeared late in 1938. In January, 1940, *Language in General Education* and *A Bibliography of Prose Fiction* were published. *The Visual Arts in General Education, The Social Studies in General Education,* and *Mathematics in General Education* are scheduled for publication in 1940. In addition, there will be a number of volumes from the Adolescent Study itself which represent outcomes of the case study method applied to some six hundred students in secondary schools and colleges. The major publications will be:

Peter Blos, *Youth in the Making: A Study in Adolescent Growth.*

Elizabeth Hellersberg, *Adolescence: A Period of Transition.*

Wilma Lloyd, *Observation and Objectivity.*

Caroline B. Zachry and Margaret E. Lighty, *Emotion and Conduct in Adolescence.*

Regents' Inquiry.—According to information secured from Luther Gulick, director of the survey, the chief contributions of the Regents' Inquiry in New York are as follows:

1. The Inquiry has set in motion a greater volume and a higher quality of general lay discussion of the entire school program in New York State than has heretofore prevailed. The League of Women Voters organized over 80 study groups to examine the Inquiry reports and to analyze local school problems in the light of the larger program. The Parent Teacher Association had over 600 similar local groups out of some 800 in the state engaged in a study of the reports.

2. The Inquiry findings served as material for discussion in hundreds of local school systems throughout the state. In some cases reports were given by individual teachers from the Inquiry documents, and in other instances whole sections of the reports were mimeographed and distributed as a basis for discussion.

3. The Inquiry served to stress the unity and integration of the whole school problem within the state, although it did throw renewed emphasis upon: (a) the elevation of rural school standards and administration; (b) the testing of the school program on the basis of its outcome measured in terms of youth; (c) the need for higher teacher standards; (d) the importance of strengthening local administration, and transforming the state department from an administrative agency into one primarily concerned with research and stimulation; (e) increasing the efficiency and economy of school administration; and (f) increasing the importance of social competence as a central educational objective.

4. The State Department of Education has been reorganized quite along the lines suggested and a Department of Research has been created directly under the Commissioner; a large amount of district school reorganization has been carried through; the Regents' Examination System is being revised extensively; both centrally and locally there is quite a new attitude with reference to education for citizenship and education for work; and nobody now talks about education, or acts about education, in the state without drawing upon the factual material and the ideas advanced by the Inquiry.

5. At least a score of the better state departments of education have utilized the reports for special conferences and staff discussions, and at least a dozen schools of education during the summer of 1939 organized special courses based on the Regents' Inquiry reports.

Research Planning in New York.—The Educational Research Association of New York State has issued the report of a special planning committee, under the chairmanship of Arvid J. Burke. The function of the committee is to formulate plans for a state research program which will show the interrelations and respective responsibilities of the State Education Department, the State Research Association, the State Teachers Association, the Elementary Principals Association, the Association of Academic Principals, the other state-wide educational organizations, the graduate schools of education, and the local schools and colleges.

The research needs of the state are outlined as follows: (1) to determine the most important problems calling for research in education at a given time; (2) to clarify the issues involved in each division of education and for defining specific problems in relation to these larger issues; (3) to utilize fully existing educational research facilities through voluntary coöperation and coördination of effort, having each agency assume the responsibilities which it can best handle; (4) to lift the level of educational research or to improve the quality of much so-called research through clarification of research concepts, improved methodology, formulation of research standards, and the dissemination of information on these among research workers and consumers of research; (5) to provide more adequate support for educational research, especially in the State Education Department, the public school systems, and the institutions of higher education; and (6) to implement research findings.

The committee states five principles for carrying out a coördinated program of educational research:

1. Research will not function best in a "line organization," designed efficiently to carry out fixed-in-advance plans. Research must be organized to promote intellectual freedom, critical attitudes, and creative effort. In planning a research program, authority and control should be reduced to a minimum, with stimulation and leadership substituted in its place.
2. The coördinating of educational research activities in the state should be attained through voluntary coöperation (on the part of individual workers and research agencies), a central body merely acting as the agency to facilitate this voluntary coördination.
3. Initiative and freedom to determine its own research program should be left to each agency.
4. The freedom of the individual in scientific research should be protected.
5. Research workers, insofar as possible, should have no responsibility for formulating or carrying out the policies which they are studying.

Florida Projects.—A 430 page volume comprises the report of a survey of the schools of Dade County, Florida, which was recently completed by the Florida State Department of Education. Edgar L. Morphet, Director of Administration and Finance of the State Department, was director of the survey. The staff was comprised of representatives of the state department of education and of the institutions of higher learning in the state. Bess Goodykoontz, Assistant Commissioner of Education, and H. F. Alves, Senior Specialist in State School Administration, served as consultants from the United States Office of Education. This report is necessarily unique in many respects, because it deals with problems which are duplicated in few areas of the United States. These problems, however, are of interest to thousands of people from all sections of the country, who come to Miami each year and whose children attend schools in the Miami area. The survey staff used this study as an opportunity to think through and further develop a philosophy of education for Florida schools to aid in solving many of the outstanding educational problems in the state, as well as to make recommendations of particular importance for Dade County. This report, which is well illustrated by pictures, charts, and maps, is distributed through the Florida State Department of Education (paper bound, $1.50).

A Southern States Work-Conference on School Administrative Problems is being held at Daytona Beach, Florida, June 12-14, inclusive. The purpose of this work-conference is to enable representatives from the participating states to work together in developing handbooks or manuals for use in their own states, having as much in common as practicable. The projects on which studies will be undertaken are: (1) continuity of service (tenure) for teachers, (2) operation and maintenance of transportation, and (3) operation and maintenance of the school plant. The advisory staff is comprised of H. F. Alves, United States Office of Education; Howard Dawson, National Education Association; and John Guy Fowlkes, University of Wisconsin. The participating staff will be comprised of members of the staff of the state departments of participating states. All southern states have made plans to participate in this significant work-conference.

Research on the School Plant.—T. C. Holy of Ohio State University calls attention to the need for an organized program of research in the school plant field, an area which involves approximately one-third of the total expenditures for public education. The School Plant Research Committee of the American Council on Education has received a grant of $12,000 for a study of school equipment specifications and standards. As illustrations of needed research, unsolved problems are listed from the field of heating and ventilation, an area which has been more extensively explored than any other in the school plant field:

1. Is open window ventilation a satisfactory method under all conditions? If not, under what conditions is it satisfactory?

2. If mechanical systems of heating and ventilating are used, how many cubic feet of air per pupil are required?

3. To what extent and under what circumstances can air be safely recirculated?

4. What are the optimum temperature and humidity conditions for health and learning?

5. Is air conditioning feasible and economical under typical school conditions?

6. What are the relationships between building design and building materials and ventilation problems?

Minnesota Studies.—The report of the University of Minnesota Committee on Educational Research for 1936-38 includes a tabulation of the number of studies conducted in each field from 1924 to 1938: administration, 6; alumni, 2; articulation, 11; examinations and measurements, 12; higher education, 6; instructional, 21; library, 2; physical education and recreation, 1; student personnel, 13; total, 74. The total expenditure for these investigations since 1931 is $107,789. The chairman of the Committee is T. R. McConnell.

Research in Australia.—The Australian Council for Educational Research is planning the sponsorship of a journal devoted to educational research. During the twelve months preceding August, 1939, the Council distributed more than 150,000 copies of standardized tests to the schools of Australia. The Carnegie Corporation of New York has continued at a reduced rate its original ten-year grant, which terminated at the end of 1939, to the extent of $75,000 to assist the Council over a three-year period, with the expectation that adequate local aid will then be forthcoming. Recent publications of the Council are: *Review of Education in Australia, 1938; Victoria's First Public Educationist,* by E. Sweetman; and *The Educational Needs of a Rural Community,* by W. C. Radford.

Guggenheim Fellowships, 1940.—Seventy-three fellowships, with stipends totaling $165,000 to assist research and creative work, have been announced by the John Simon Guggenheim Memorial Foundation. The Foundation, established in 1925 by former United States Senator and Mrs. Simon Guggenheim as a memorial to a son, to assist original work by scholars and artists, in the past fifteen years has granted 913 fellowships and $2,082,000 to assist its fellows to carry on their work. On account of the wars in Europe and the Far East all fellows, except one whose plans will take him to the Near East, will work in the Western Hemisphere this year. Twelve will go to Latin America, and the rest will conduct their studies in the United States and in Canada. Among those who will work in the United States are six Canadians appointed under a recent extension of the Foundation's fellowships to Canada. The annual Guggenheim Fellowships for Latin American scholars and artists, another part of the Foundation's plans for granting fellowships in this hemisphere, will be awarded in June.

Rockefeller Foundation, 1939.—Grants made by the Rockefeller Foundation in 1939 amounted in round figures to $9,500,000. Of this sum, $8,000,000 was distributed in approximately equal amounts of $2,000,000 each to work in the four fields of public health, medical, social, and natural sciences. One million dollars was appropriated for activities in the humanities, and the better part of a half million toward the program of rural reconstruction in China. The income of the Foundation during 1939 was $6,627,450. The program of the Foundation in 1939 involved coöperative effort in 47 countries. The ratio between the amount of money that went to foreign countries and the amount spent on work in the United States was the same as last year, i.e., approximately 25 percent in foreign countries as against 75 percent in the United States.

Evaluation of Business Education.—A Work Committee, composed of representative educators within and outside the field of business education, has been organized in connection with the Annual Conferences on Business Education held at the University of Chicago. The responsibility of this committee is to initiate procedures for setting standards which will be of assistance to school administrators, teachers, and accrediting agencies in measuring business curriculums and guidance programs.

Last year the committee submitted a tentative report of a series of criteria and standards for appraising secondary-school economic and business education. A number of tentative and general statements and recommendations were formulated with respect to the following areas: (1) purposes and philosophy of business education, (2) the curriculum, (3) the teaching staff, (4) facilities, (5) guidance program, and (6) adjustment to economic and community situations.

The committee this year is at work refining the rating scale on the basis of criticisms by school administrators and curriculum specialists. A second major task is to formulate a testing program with emphasis on the general economic and business background phases of education. This testing program is to be used as a means of measuring the outcome of the business subjects, and also to make some comparison between the business students and the non-business students. A third major phase of the Work Committee's activity for the present year is to apply the revised criteria in terms of an actual school situation. The Fort Wayne public schools are being

used for the purpose of making an actual appraisal of school practices.

A copy of the tentative rating scale may be obtained from the Work Committee, School of Business, University of Chicago.

Rural Education.—A new Committee on Rural Education with offices at 600 South Michigan Avenue, Chicago, has recently been appointed by the American Country Life Association. This committee held its first meeting on March 3-4, 1939, with three subsequent meetings, and is now fully organized under the chairmanship of Roscoe Pulliam, President of the State Teachers College, Carbondale, Illinois. Iman E. Schatzmann of Iowa background, but formerly educational investigator for the International Bureau of Education and research assistant for the International Institute of Agriculture, Rome, has been appointed executive secretary.

While generally interested in rural education as a whole, and in all parts of the country, this new committee is giving initial attention to the Mississippi Valley and to the cultural and spiritual values of life on the land. It regards education as a life-long, all-inclusive process, involving the entire population and many varied agencies in addition to the school. It views education as directly related to the vital interests and needs of the people and as involving initiative, participation, and support on the part of the local community.

In formulating its program for the improvement of rural education, the committee is taking a strong stand on federal aid for rural schools as essential in securing an adequate financial basis for the education of rural children. Even more immediate is its concern for effective, life-related teaching in rural schools. To this end the committee is focusing first attention upon rural school supervision and teacher education, advocating practical demonstrations in both fields. It desires, also, to see special attention given to the unemployed and out-of-school youth of rural communities and to the numerous problems of the rural high school serving adolescent youth in school.

In the implementation of its program the committee is endeavoring to discover the best achievements and practices of rural education both in this country and abroad. These materials will be assembled and disseminated to all interested individuals and agencies. The committee is also encouraging the organization of regional, state, and local commissions or councils on rural education, and is desirous of coöperating with other agencies in promoting conferences, institutes, and demonstrations in this field. Its greatest hope, however, is centered in the quality, idealism, and skill of American rural teachers. Toward the realization of these standards, it is attempting to secure funds for scholarships and the advanced training of superior young people who will promise to make rural education their life work. The committee also is working for improvement in both the pre-service and in-service education of rural teachers, and is trying to stimulate increased activity on the part of all teacher-training institutions having farm constituencies. One demonstration in professional rural school supervision coöperatively developed between a state teachers college and a local county has been tentatively approved.

For the present the committee is being supported by a grant from the Farm Foundation of Chicago.

PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

MAY 27 1940

MAY, 1940

Journal of Educational Research

Registered June 4, 1919

CONTENTS

		PAGE
Evaluating Evaluation	*Pedro T. Orata*	641
A Study of the Extent of Measurement of Differential Objectives of Instruction	*T. R. McConnell*	662
The Wisconsin Study of Teaching Ability	*A. S. Barr*	671
An Effort to Extend the Measurement of the Results of Schooling into the Social and Economic Fields	*Harold F. Clark* and *Maurice F. Seay*	685
Use of Interest Inventories for Personality Study	*G. V. Shevtakov* and *Jean Friedberg*	692
Results from Factor Analysis	*John M. Stalnaker*	698
Selective Application of Thurstone Test	*Harold S. Tuttle*	705
Editorial: The Growing Complexity of Measurement	*A. S. Barr*	710
Reviews		712
Research Abstracts and Bibliographies		717
Research News and Communications		721

THE JOURNAL OF EDUCATIONAL RESEARCH is indexed in the Education Index

---

Copyright 1940, by Public School Publishing Co.

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879

JOURNAL OF EDUCATIONAL RESEARCH

EDITORIAL BOARD

A. S. BARR, *Chairman, Professor of Education, University of Wisconsin, Madison, Wisconsin.*

E. J. ASHBAUGH, *Dean, College of Education, Miami University, Oxford, Ohio.* Administration, Supervision, and Teacher Training.

HARRY J. BAKER, *Director, Psychological Clinic, Detroit Public Schools, Detroit, Michigan.* Child Welfare, Guidance, and Development.

LEO J. BRUECKNER, *Professor of Education, University of Minnesota, Minneapolis, Minnesota.* Elementary Education.

CARTER V. GOOD, *Professor of Education, University of Cincinnati, Cincinnati, Ohio.* Teaching, Learning, and Individual Differences.

KAI JENSEN, *Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.* Preschool Education.

A. K. LOOMIS, *c/o Board of Education, Shaker Heights, Cleveland, Ohio.* Curriculum Construction.

DOUGLAS E. SCATES, *Associate Professor of Education, Duke University, Durham, N. C.* Field Research.

PERCIVAL M. SYMONDS, *Professor of Education, Teachers College, Columbia University, New York City.* Mental Hygiene and School Instruction.

CLIFFORD WOODY, *Director, Bureau of Educational Reference and Research and Professor of Education, University of Michigan, Ann Arbor, Michigan.* Measurements, Statistics, and Research.

To save time, manuscripts may be sent directly to members of the editorial board as indicated above.

ASSOCIATE EDITORS

FRED C. AYER, *Professor of Educational Administration, University of Texas, Austin, Texas.*

EMMETT A. BETTS, *Research Professor and Director of the Reading Clinic, Pennsylvania State College, State College, Pennsylvania.*

PHILIP A. BOYER, *Director of Educational Research School District of Philadelphia, Philadelphia, Pennsylvania.*

FOWLER D. BROOKS, *Head of Departments of Education and Psychology, DePauw University, Greencastle, Indiana.*

WILLIAM A. BROWNELL, *Professor of Educational Psychology, Duke University, Durham, North Carolina.*

B. R. BUCKINGHAM, *Directing Editor of Elementary School Textbooks, Ginn and Company, Boston, Massachusetts.*

W. W. CHARTERS, *Director, Bureau of Educational Research, and Professor of Education, Ohio State University, Columbus, Ohio.*

VIRGIL E. DICKSON, *Superintendent of Schools, Berkeley Public Schools, Berkeley, California.*

EDGAR A. DOLL, *Director, Department of Research, Training School, Vineland, New Jersey.*

WILLIAM S. GRAY, *Executive Secretary of the Committee on the Preparation of Teachers, University of Chicago, Chicago, Illinois.*

CHARLES E. GREENE, *Superintendent of Schools, Denver Public Schools, Denver, Colorado.*

V. A. C. HENMON, *Chairman Department of Psychology and Professor of Psychology, University of Wisconsin, Madison, Wisconsin.*

KARL J. HOLZINGER, *Professor of Education, University of Chicago, Chicago, Illinois.*

ERNEST HORN, *Director, University Elementary School and Professor of Education, State University of Iowa, Iowa City, Iowa.*

EARL HUDELSON, *Dean, College of Education and Professor of Education, West Virginia University, Morgantown, West Virginia.*

FREDERICK J. KELLY, *Chief, Division of Higher Education, Office of Education, Federal Security Agency, Washington, D. C.*

ELISE H. MARTENS, *Senior Specialist in the Education of Exceptional Children, Office of Education, Federal Security Agency, Washington, D. C.*

WILLIAM A. McCALL, *Professor of Education, Teachers College, Columbia University, New York City.*

W. J. OSBORN, *Professor of Education, University of Washington, Seattle, Washington.*

JESSE B. SEARS, *Professor of Education, Stanford University, California.*

MANDEL SHERMAN, *Associate Professor of Educational Psychology, University of Chicago, Chicago, Illinois.*

FRANCIS T. SPAULDING, *Professor of Education, Graduate School of Education, Harvard University, Cambridge, Massachusetts.*

JOHN L. STENQUIST, *Director, Bureau of Educational Research, Department of Education, Baltimore, Maryland; and Lecturer, University of Maryland.*

W. W. THEISEN, *Assistant Superintendent of Schools, Milwaukee, Wisconsin.*

M. R. TRABUE, *Dean, School of Education, Pennsylvania State College, State College, Pennsylvania.*

CARLETON W. WASHBURNE, *Superintendent of Schools, Winnetka Public Schools, Winnetka, Illinois.*

ELIZABETH L. WOODS, *Supervisor, Educational Research and Guidance, Los Angeles City Schools, Los Angeles, California.*

J. WAYNE WRIGHTSTONE, *Assistant Director, Bureau of Reference, Research, and Statistics, Board of Education, New York City.*

G. A. YOAKAM, *Professor of Education and Director, Courses in Elementary Education, University of Pittsburgh, Pittsburgh, Pennsylvania.*

The JOURNAL OF EDUCATIONAL RESEARCH is published at Bloomington, Illinois, monthly, except June, July, and August by the Public School Publishing Company. The subscription price is $3.50; $3.60 in Canada, and $3.90 in other foreign countries.

Subscriptions and all communications about reprints or other business matters should be addressed to the PUBLIC SCHOOL PUBLISHING COMPANY, at Bloomington, Illinois.

Address all communications on editorial matters to A. S. Barr, Department of Education, University of Wisconsin, Madison, Wisconsin.

• RETRIEVING THE RETARDED READER

• with special emphasis on remedial teaching of vocabulary

• by Jack W. Birch
Instructor, Orthogenic Retarded Pupils
Burnham, Pennsylvania, Public Schools

Birch says, "The problem of the retarded reader is one of *school organization* in that it deals with a group of children whose needs have not been met by the mechanism of the school in which they participate. It is a problem of the *child* in that he realizes he is not able to participate with maximum satisfaction in the educative process set up for him. It is a problem of the *teacher* in that he realizes his responsibility to educate the child cannot be met adequately. It is a problem of *those who endeavor to devise functional materials and methods of instruction,* and it is a problem of *those who do research in education and psychology.*"

Birch's new booklet on **"Retrieving the Retarded Reader"** is packed with sound and practical suggestions for work with poor readers. In simple, non-technical language, it discusses the problem of the retarded reader and emphasizes the remedial teaching of *vocabulary* on the preprimer, primer, elementary, and secondary school levels.

So that each of your teachers may be supplied with a copy of this new pamphlet, we have priced **"Retrieving the Retarded Reader"** at 30¢ a single copy, postpaid, or 25¢ each, plus postage, in quantities of ten or more copies.

Order from

PUBLIC SCHOOL PUBLISHING COMPANY
BLOOMINGTON, ILLINOIS

USE TESTS VALID FOR THE SCHOOL COURSE AND CLASS INSTRUCTION

With a measuring scale which correlates both aptitude and attainment in school work to chronological age, our standardized tests have been so adjusted as to measure the success of each pupil under the conditions of his own abilities and of his interest and effort as directed under class and instructional conditions.

Three basic principles for a definite measure of the success of school instruction are covered by the *Public School Correlated Attainment Scales*. These principles are:

First, all the tests used are validated for the teaching and learning conditions in the class, as well as with reference to the textbooks.

Second, the three aptitude tests are fully validated for pupils in grades 4 to 9 as measures of the aptitude needed for the studies covered by the attainment tests.

Third, an important source of school success is found in the instructional equipment as it relates to the interest and effort given by each pupil. Reliability for this situation is secured by the adjustment of the order of difficulty for each question in each test to the aptitude ages found in the class.

By all of these tests, each pupil's work is measured in terms of the correlated scale, which scale is standardized for the present progress of the median pupil through the elementary grades. Each pupil in his work in each subject has the attainment recorded on this standard measure of school progress.

Schools which have used these *Public School Correlated Attainment Scales* and have made the detailed analysis which the record chart provides are enthusiastic over the value of the scales for locating specific instructional needs of each pupil and of the class.

Three forms are provided. Use the revised *Form A* for your first testing. Ask for your sample set to be sent on approval.

PUBLIC SCHOOL PUBLISHING COMPANY

BLOOMINGTON, ILLINOIS